Spring 1-1-2012

Non-Traditional Approaches to Interrogating Genome-Wide SNP Data

Daniel Patrick Howrigan
University of Colorado at Boulder, daniel.howrigan@gmail.com

Non-Traditional Approaches to Interrogating Genome-Wide SNP Data

by

Daniel P. Howrigan

B.A., University of California at Santa Barbara, 2002

M.A., California State University at Long Beach, 2008

M.A., University of Colorado at Boulder, 2011

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

Of the requirement for the degree of

Doctor of Philosophy

Department of Psychology

2012

This thesis entitled:

Non-Traditional Approaches to Interrogating Genome-Wide SNP Data

Written by Daniel Howrigan

Has been approved for the Department of Psychology

_____

Matthew McQueen, Sc.D.

_____

Matthew Keller, Ph.D.

_____

Marissa Ehringer, Ph.D.

_____

Gregory Carey, Ph.D.

_____

Matthew Jones, Ph.D.

Date: July 11th, 2012

The signatories have examined the final copy of this thesis, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above-mentioned discipline.

Howrigan, Daniel (Ph.D. Psychology)

Non-Traditional Approaches to Interrogating Genome-Wide SNP Data

Thesis directed by Professor Matthew B. McQueen and Matthew C. Keller

The field of human molecular genetics has undergone a substantial technological transformation in the past decade, allowing researchers to identify and analyze genetic variation across the human genome with unprecedented depth and precision. A central goal in utilizing this technological advancement is to discover the genetic variation that underlies complex heritable traits and disorders. In recent years, much of the focus has been on large-scale genome-wide association studies (GWAS) in an attempt to identify effects of common single nucleotide polymorphisms (SNPs) on a phenotype. Progress in this arena, however, has been limited, as validated findings for most phenotypes represent only a small fraction of the variance attributed to genetics known from family and twin studies, leaving a large proportion of heritability to be explained. My research is in large part motivated by the issues surrounding GWAS, as additional methodological techniques and population genetic theory can help explain phenotypic variance unaccounted for by the traditional genetic association design. In my first study, I look at a case/control sample of bipolar disorder, examining how prior information from linkage studies can inform GWAS signals. The primary aim is to ease the burden of multiple-testing correction applied in GWAS using empirically informed weighting to tease out true signals supported by prior genetic evidence. In my second study, I determine the best practices to detect signatures of distant inbreeding, via runs of homozygosity, in genome-wide SNP data. The motivation for studying this phenomenon is due to the extensive evidence of inbreeding depression on fitness related traits, where the effects of recessive or partially recessive alleles are expressed. Using an

extensive simulation design, I test multiple programs to determine the optimal method to identify runs of homozygosity caused by distant inbreeding. In my third and final study, I apply my work from the second study to a comprehensive dataset of cognitive measures to understand the extent to which distant inbreeding affects variation in general cognitive ability.

Dedication

To my parents - for their love, patience, and unwavering support throughout my academic training

Acknowledgements

I would like to thank several people for their help in completing this dissertation. First and foremost my advisors Matthew Keller and Matthew McQueen for guidance, support, and creating a wonderful environment for my learning and growing in a new field.

I would also like to thank fellow IBG faculty for inspiring my learning - most notably classes with Greg Carey, Tom Johnson, Marissa Ehringer, Don Cooper, and Soo Rhee that tested my knowledge of behavioral and molecular genetics, and fostered a greater understanding of the variety of approaches to gain insight into the genetic basis of behavior.

I would also like to acknowledge my entire dissertation committee for their suggestions and helpful feedback.

Finally, I would also like to acknowledge my fellow graduate students and lab members who have taught me, inspired me, and helped me become a more successful member of the genetics community.

This work was primarily supported by the following project grants: HD007289, MH016880, MH085812-01A1

Contents

Tables

Figures

Chapter 1

Introduction

In the field of human molecular genetics, the past decade has witnessed a rapid transformation in terms of technological advancement and feasibility to detect and analyze molecular genetic variation at a large scale. Not long after the initial sequencing of the first human genomes in 2001 (Lander et al., 2001; Venter et al., 2001), the identification of genetic variants at the base pair level became possible. Less than a decade later, genome-wide association studies (GWAS) using databases up into the tens of thousands of individuals, each one genotyped with up to a million single nucleotide polymorphisms (SNPs) across the genome, have been gathered to identify the genetic architecture of complex traits and disorders such as Schizophrenia (Purcell et al., 2009), Type 2 Diabetes (Zeggini et al., 2008), and height (Weedon et al., 2008), just to name a few. However, despite the technological advances and large-scale effort in this arena, there is still much to be discovered in understanding the genetic underpinnings of most complex traits and diseases (Hardy & Singleton, 2009; Manolio et al., 2009). The research I present in this dissertation centers on genome-wide SNP data, and the methods I use to analyze the data are closely related to central issues that surround GWAS. On that account, my introduction briefly reviews the progress of molecular genetic methods in studying human phenotypes, examining both general findings and prominent issues underlying genome-wide analysis. This leads to my motivation for using non-traditional approaches to analyze genome-wide SNP data, and brief description of the methods I adopt for my dissertation research.

1.1 Human Molecular Genetic Methods

Measurable forms of molecular genetic variation have been around for well over a century, as observations of chromosomes through karyotyping were discovered as early as 1842 (Sturtevant, 2001), marking the start of cytogenetic analysis. For much of the early 20th century, genetic linkage analysis used recombinant strains to measure multiple traits and infer the genetic distance of trait loci on chromosomes. However, this process of linkage mapping was not able to isolate the physical position of trait loci, and it wasn't until molecular techniques such as restriction enzymes and gel electrophoresis that physical mapping of the genome became possible. With respect to analyzing human genetic variation, one of the earliest widespread applications of identifying physical stretches of DNA sequence was through restriction fragment length polymorphisms (RFLPs), where restriction enzymes would cut DNA at homologous sites (Southern, 1975). As a result, differing length sequences that remained were treated as alleles in that chromosomal stretch. This process was updated with the incorporation of oligonucleotides (small stretches of DNA) and polymerase chain reaction, where instead of DNA being cut at only at sites that match restriction enzymes, a targeted stretch of DNA could be probed and amplified, allowing access to any genetic region were the designed probe matched the DNA sequence (Saiki et al., 1985). Soon variable number tandem repeat (VNTR) polymorphisms were preferred over RFLPs, as their abundance allowed for finer-scale mapping of genomic regions. Particularly for trait and disease mapping, short tandem repeats (often called STRs or microsatellites) where only a few base pairs were repeated, often dozens of times, were targeted as allelic markers. STRs occurred on average once per 10 kilobases of DNA, and soon genome-wide STR marker sets were being used in linkage studies with up to a thousand markers (Altmüller, Palmer, Fischer, Scherb, & Wjst, 2001).

Nevertheless, not long after microsatellites became commonplace, the results of the human genome sequencing projects around the turn of the 21st century identified single nucleotide polymorphisms (SNPs), where a single base pair varied in its nucleotide base between individuals. Because SNPs occurred at a much higher density than VNTRs and were abundant in and around genes, they soon became the predominant substrate to catalog genetic variation. The advance in recent years has been the accelerating pace and number with which SNPs can be genotyped, from thousands to over 1 million SNPs in a genome-wide array. The pace of technological improvement is only getting faster and cheaper, as other recent advancements have included copy number variation (CNV; Sebat et al., 2004; Sharp et al., 2005), SNP imputation via patterns of linkage disequilibrium (LD; Gibbs et al., 2003), and most recently whole-genome sequencing (Thousand Genomes Consortium, 2010).

1.2 Human Molecular Genetic Analysis

As methods of identifying and mapping genetic variation have evolved, so too have the methods to analyze their effects on phenotypic variation. Drawing from the insights of Mendelian inheritance and Morgan's work on fruit flies, genetic linkage mapping began as the predominant paradigm to infer causative loci for a trait. Linkage follows from the rules of meiosis and recombination, where genetic linkage refers to statistical propensity of alleles on a given chromosome to occur together because they haven't been broken up by meiotic recombination. Because recombination occurs at a low (although stable) rate per chromosome during meiosis, the distance between two loci greatly affects how strongly they are linked. Using this logic, linkage mapping was developed as a means to infer the chromosomal location of causal loci via transmission from parent to offspring. For human genetics, this meant using family-based studies and pedigree information, where parent to offspring transmission of alleles

could highlight the co-transmission of alleles and phenotype status (Lander & Green, 1987). RFLPs and VNTRs, with their numerous alleles (sometimes in the dozens), were good sources of genetic variation for family-based linkage analysis as each parent often carried a different set of alleles. Linkage work spawned the discovery of many mendelian disorders, as well as discovering the genes that predispose individuals to Huntington's disease (Gusella et al., 1983), phenylketonuria (DiLella, Marvi, Brayton, & Woo, 1987), cystic fibrosis (Kerem et al., 1989), and a handful of other genetic disorders. Despite its successes, linkage designs suffered from low power to detect variants of modest effect, as markers had to be strongly linked to the causal variant and required a large number of informative families to be detected (Altmüller et al., 2001). Along with the introduction of SNPs, linkage gave way to association mapping, where the allele itself, not its transmission from parent to offspring, became the variable of interest. SNP association still incorporated the concepts of genetic linkage, as genotyped SNPs in linkage disequilibrium (LD) with a nearby causal SNP would segregate together, allowing the statistical signal from the genotyped SNP to be representative of the causal SNP in the population. As SNPs offered much denser coverage of genetic variation, association mapping offered the additional statistical power and ability to use population-based samples rather than family samples (Balding, 2006).

Throughout the transformation of genetic methodology, there has been a relative shift from studying biologically plausible candidate genes towards genome-wide interrogation. While debates over which method was the most effective use of research dollars have waned, mainly because costs to run genetic analyses have lowered considerably in recent years, arguments still persist on both sides. Further spurring this debate is the slow realization that outside of the handful of big findings in human genetics, most genetic effects are very difficult to find

regardless of the methodology used (Hardy & Singleton, 2009). More specifically, during the early days of linkage, monogenic disorders of large effect were readily identified, as their genetic signals were unmistakable. Once the 'low-hanging fruit' of large genetic effects had been picked, the search for causal genes or variants among phenotypes with a complex genetic etiology proved much more challenging. While proponents of candidate gene work criticized the large costs of genome-wide association studies (GWAS) at the outset, one of its main attractions was the ability to vastly speed up the process of susceptibility gene discovery and validation, as high-density SNP arrays were able to cover the majority of common genetic variation in the genome. In addition, GWAS has offered an agnostic criterion for identifying association signals, whereas prior candidate gene approaches may have been biased towards false positive results and publication bias (Hirschhorn, Lohmueller, Byrne, & Hirschhorn, 2002). From a theoretical standpoint, the central model underlying GWAS was the 'common disease/common variant' hypothesis, asserting that complex traits and diseases that are heritable but do not follow Mendelian patterns of inheritance must be attributable in some respects to common genetic variation (> 1-5%) shared in the population (Cargill et al., 1999; Reich & Lander, 2001). While disease etiology using the common disease/common variant hypothesis was by no means the consensus among geneticists (e.g. Pritchard, 2001), the technological advance of SNP genotyping was paramount in shifting the field toward genome-wide analyses.

1.3 Summary of GWAS Results

Since the first GWA studies in the mid 2000's, over 1,200 loci have been associated with over 150 traits and diseases (Hindorff et al., 2009; Lander, 2011; and a summary of results can be found at http://www.genome.gov/gwastudies/). In the genetic epidemiology literature, GWAS has led to the discovery of moderate to large effects in a handful of disorders such as macular

degeneration (Maller et al., 2006), Crohn's disease (Barrett et al., 2008), type 2 diabetes (Zeggini et al., 2008), and others. The majority of replicable findings in GWAS, however, have found relatively small effects at best, with associated SNPs often explaining less than 1% of the phenotypic variance of the trait being studied. This has certainly been the case with psychiatric disorders, where large scale GWAS on highly heritable disorders such as schizophrenia and bipolar disorder have only discovered a handful of associated SNPs, each with very small effect sizes (Ferreira et al., 2008; Purcell et al., 2009; Shi et al., 2009b). To confuse matters even more, the analysis of a large number of SNPs simultaneously requires a strict correction for multiple testing, possibly missing true SNP associations of even smaller effect in order to avoid false positives. The relatively small amount of phenotypic variance explained by GWAS for many well established heritable traits and diseases has left what some have dubbed the 'missing heritability', and has led to debates about the success and future of GWAS studies, as the associations found were much smaller than anticipated, and the investment in larger samples to detect increasingly smaller effects must compete with the renewed interest in rare genetic variation detectable using newer sequencing technologies. Suffice to say, the ‘common variant/common disease’ model has not held up strongly for most traits, and alternative models are currently being debated to influence the direction of future research.

1.3.1 Alternative Methods and Future Directions

Although a fair number of hypotheses have been offered to explain the missing heritability, from epigenetic mechanisms (Petronis, 2010) to questioning the merit of classical heritability estimates (Zuk, Hechter, Sunyaev, & Lander, 2012), two prominent explanations have surfaced: 1) The heritability of the trait is driven primarily not by common SNPs, but rather rare variants or copy number variants not tagged by common SNPs, and 2) The full effect of

common SNPs are too small to be detected individually but on aggregate explain a substantial portion of phenotypic variance. Despite proposing quite different genetic etiologies, these are not mutually exclusive hypotheses, as a phenotype can be highly polygenic and still be substantially affected by a rare mutation at a single locus. The current move towards whole-genome sequencing seeks to address the influence of rare variation on traits, as these variants will not effectively be picked up by common SNP arrays. On the other hand, methods aggregating SNP signals (Purcell et al., 2009) or estimating phenotypic variance using all SNPs (Yang et al., 2010a) have shown that traits with otherwise few detected loci show signs of highly polygenic inheritance.

1.4 Summary of Dissertation Methods

In my dissertation research, I examine and apply two methods outside of the traditional SNP association approach, one being the use of prior linkage information to inform association signals of bipolar disorder, and the other examining the best practices to measure genome-wide autozygosity and applying this work towards measures of cognitive ability. Both methods have implications for addressing the 'missing heritability', as prior information can help tease out the real effects of signals not captured by traditional GWAS, whereas autozygosity measurement and mapping measures the role of dominance variation in the genetic architecture of a trait. With respect to common genetic variation, SNP-based heritability methods have shown that many SNPs contribute to trait variance in height not detected by GWA studies (Yang et al., 2010a), and the use of prior linkage information can help identify specific association signals that have been missed due to strict correction for statistical significance (Manolio et al., 2009; Roeder, Bacanu, Wasserman, & Devlin, 2006). With respect to rare genetic variation, measuring autozygosity can highlight the contribution of recessive or partially recessive alleles, both

common and rare, on a phenotype. Autozygosity is the genetic measurement of inbreeding, where an autozygous segment in an individual is identical by descent. By identifying autozygous segments in SNP data, one can infer that the DNA within that segment is also identical. Because deleterious recessive alleles are not as effectively removed from the population as additive alleles, and that inbreeding depression is a widely supported phenomenon in the animal kingdom (DeRose & Roff, 1999a), it is reasonable to presume that deleterious recessive alleles, particularly those a low frequencies in the population, are contributing to phenotypic variance.

1.4.1 Chapter Outline

In my second chapter, I examine how prior linkage information can be used to inform GWA signals in a case/control dataset of Bipolar Disorder, a psychological disorder that has had only a few successful replicable GWAS signals. Using prior linkage data combined from eleven separate studies (McQueen et al., 2005), I apply a Weighted False Discovery Approach (wFDR; Roeder et al., 2006) to a genome-wide scan of bipolar disorder. In my third chapter, I use an extensive simulation paradigm to understand the best methods to detect genome-wide autozygosity. In particular, I test three separate programs designed to detect autozygosity on simulated genetic data, determining the most optimal program and parameters within that program for capturing autozygosity. Finally, in my fourth chapter, I apply my work on autozygosity to a comprehensive dataset of cognitive ability; testing the prediction that increased autozygosity burden leads to lowers scores on measures of intelligence.

Chapter 2

Using Linkage Information to Weight a Genome-Wide Association of Bipolar Disorder

2.1 Background

Bipolar disorder (BPD) is a debilitating mental disorder that is common in the population (1-4% depending on the specific BPD classification (Merikangas et al., 2007)), yet has a complex etiology and life course across individuals. Family, twin, and adoption studies on BPD have convincingly shown that there is a substantial heritable component (Edvardsen et al., 2008; Smoller & Finn, 2003), leading researchers to search for susceptibility genes, both in candidate regions and across the genome, that predispose individuals to BPD. To date, despite the high heritability of BPD, the discovery of susceptibility genes has been a challenging endeavor.

Hypothesis-driven candidate gene approaches, most of which target genes involved in neurotransmitter systems, have largely been inconclusive with many initial findings failing to replicate. Some genes that have shown replication or significance in meta-analyses are the serotonin transporter (Cho et al., 2005; Lasky-Su, Faraone, Glatt, & Tsuang, 2005), brain-derived neurotrophic factor (Kremeyer et al., 2006; Neves-Pereira et al., 2002; Sklar et al., 2002), d-amino acid oxidase activator (Detera-Wadleigh & McMahon, 2006), monoamine oxidase A (Muller et al., 2007), and a gene that codes for 5,10-methylenetetrahydrofolate reductase (Gilbody, Lewis, & Lightfoot, 2007). Unfortunately, none of these replicated findings have shown a large genetic effect, leaving much of the genetic variance in BPD to be explained.

2.2 Genome-Wide Approaches to Bipolar Disorder

In contrast to candidate gene approaches, genome-wide linkage and association

approaches have taken an agnostic approach to finding susceptibility genes for BPD. Family-based linkage approaches on the whole have not consistently implicated a single region (McGough et al., 2008), but the most comprehensive meta-analysis on BPD has implicated regions of chromosome 6q for Bipolar I, and 8q for Bipolar I and II (McQueen et al., 2005). More recently, population-based genome-wide association studies (GWAS) on BPD, which are able to detect smaller genetic effects, have identified associated SNPs on chromosome 16p12 (WTCCC, 2007), the diacylglycerol kinase eta (DGKH) gene on chromosome 13 (Baum et al., 2008), and the myosin 5B (MYO5B) gene on chromosome 18 (Sklar et al., 2008). These three large-scale GWAS studies were then pooled together, finding a significant association in the ankyrin G (ANK3) gene on chromosome 10 and a replicated suggestive association signal in the alpha 1C subunit of the L-type voltage-gated calcium channel (CACNA1C) gene on chromosome 12, although none of the prior top signals were identified in the pooled analysis (Ferreira et al., 2008). Finally, a SNP in the zinc-finger protein 804A (ZNF804A) on chromosome 2 has shown association in both schizophrenia and BPD datasets, while a common polygenic approach compromising large clusters of SNPs has shown concordance in effects on schizophrenia and BPD (Williams et al., 2010), implicating shared genetic liability. Despite these large-scale efforts, the findings still represent a small proportion of the genetic variance in BPD (no top signals with an odds ratio greater than 1.6), suggesting a complex genetic etiology compromised of multiple genes with no single genetic risk factor being a sufficient cause of BPD.

2.3 Issues of Multiple-Testing Correction

One of the primary issues surrounding genome-wide analysis is the amount of multiple testing that arises from analyzing hundreds of thousands of SNPs. Although the use of

Bonferroni correction for multiple testing limits the possibility of making type-I errors, by definition it also raises the probability of committing type-II errors, possibly diminishing the chances of detecting true signals of association. This has prompted statistical methods that utilize prior information to guide association scans or assist in prioritizing genetic regions in follow-up studies (e.g. Fan et al., 2010). In the current study, we apply a weighted false discovery rate procedure (wFDR; Roeder et al., 2006), taking prior linkage information derived from a genome-wide linkage meta-analysis (McQueen et al., 2005) to variably weight association signals from a GWAS drawn from the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD; Sklar et al., 2008). By incorporating meta-analytical prior information, the current technique lends an increase in detection power through an integrated empirical approach to genome-wide analysis.

2.4 Methods

2.4.1 GWAS Methods

2.4.1.1 Case-Control GWAS Sample

Our case sample consisted of 955 Caucasian bipolar I subjects drawn from the genetic repository of the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). The STEP-BD sample is a longitudinal cohort drawn from the United States examining the effect of treatments in the course of BPD (Sachs et al., 2003). All subjects were diagnosed for bipolar I on the Affective Disorders Evaluation and the Mini-International Neuropsychiatric Interview. Our control sample consisted of 1,498 US Caucasian subjects drawn from the NIMH genetics initiative through the NIMH Center for Collaborative Studies (http://zork.wustl.edu/nimh). Of the controls, 454 came from anonymous blood cord donors and were phenotypically unscreened

(Mansour et al., 2005). The remaining 1,044 controls completed an online self-administered psychiatric screen, and qualified as controls if they reported no history of schizophrenia, schizoaffective disorder, auditory hallucinations, delusions, and bipolar disorder (see Sklar et al. 2008 for further information).

2.4.1.2 SNP Selection

Genotyping for the STEP-BD cases and NIMH controls were all performed on the Affymetrix GeneChip Human mapping 500K chipset. Quality control procedures and subsequent association analysis on the genotyped SNPs were all performed using PLINK (Purcell et al., 2007). Our quality control procedures, which involve individual exclusion, SNP exclusion, and population stratification, follow that described in Sklar et al. (2008). We include the additional step of removing all SNPs with minor allele frequency (MAF) < .05 and all SNPs on the X chromosome, as the linkage data only covers the 22 autosomes. After exclusion, 342,191 SNPs were retained for further analysis.

2.4.1.3 GWAS Analysis

We initially performed a genome-wide association analysis on the 342,191 SNPs. We chose the Cochrane-Armitage trend test to generate the primary nominal, unweighted association statistics. Figure 2.1 shows the unweighted, unadjusted *p*-values across the 22 autosomes. These results are consistent with the original publication of these data (Sklar et al., 2008). The *p*-values from this analysis were then weighted according to linkage evidence in that particular region (see below).

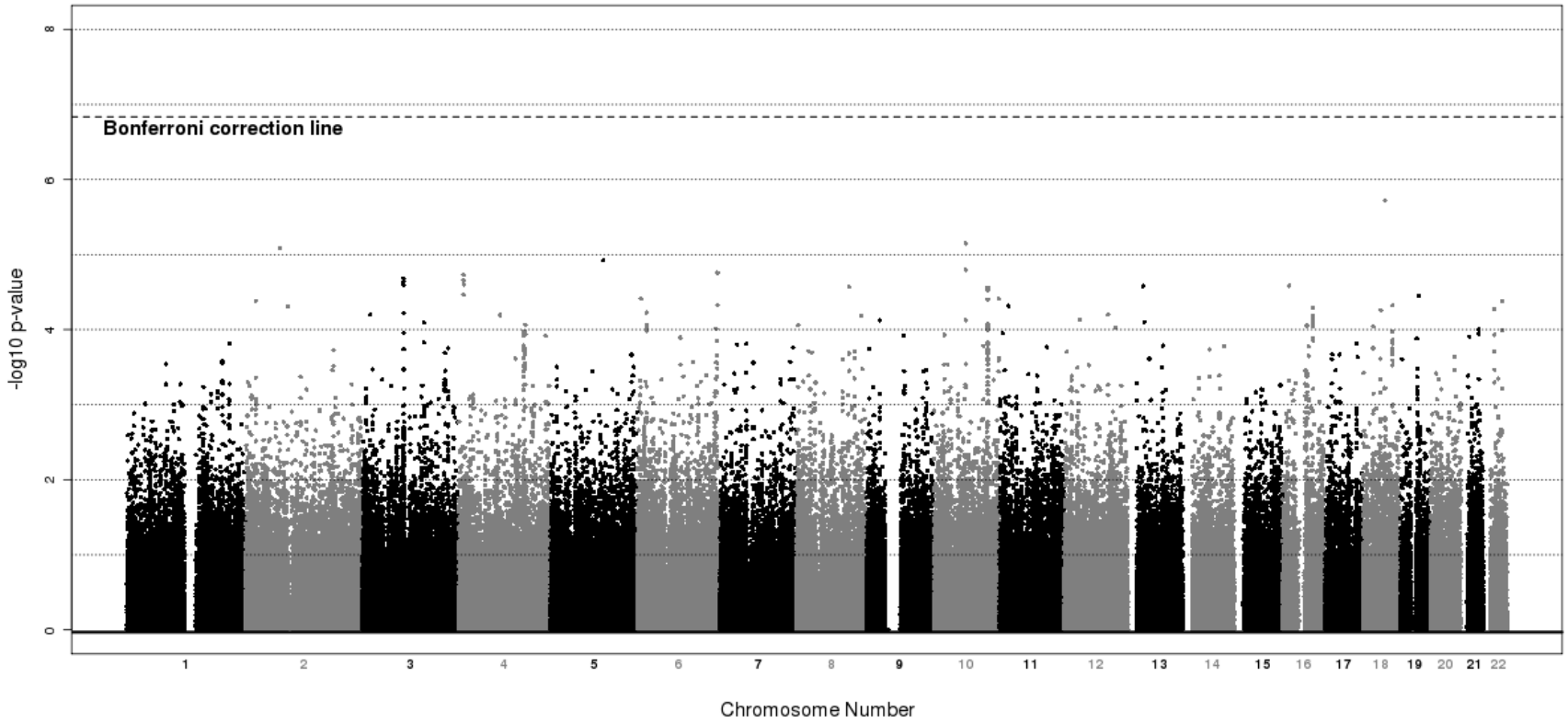

Figure 2.1 Cochran-Armitage Association Results of the STEP-BD Cohort (342,191 SNPs). The dashed line represents the genome-wide significant *p*-value threshold after Bonferroni correction.

2.4.2 Linkage Analysis Methods

For a detailed description of the combined linkage analysis samples, methods, results and conclusions, please see McQueen et al. (2005). Briefly, a genome-wide linkage trace was conducted on a combined sample consisting of eleven BPD linkage studies amounting to a total of 5,179 individuals from 1,067 families (McQueen et al., 2005). The combined sample was relatively homogenous with the majority of individuals being of Caucasian descent. Linkage statistics were generated using the affected relative pair methodology implemented in MERLIN (Abecasis, Cherny, Cookson, & Cardon, 2002) at 1 centimorgan (cM) intervals across the 22 autosomes. In the original combined analyses, two different affection status models were used - bipolar I as well as bipolar I and II. To maintain consistency with the case-control STEP-BD samples, only the linkage statistics from the bipolar I analysis ("narrow" definition) were used. Using MERLIN's implementation of the Whittemore and Halpern (Whittemore & Halpern, 1994) algorithm to test for allele sharing across all affected individuals, we generated nonparametric LOD scores via the Kong and Cox (Kong & Cox, 1997) linear model. For the purposes of this study, we used the corresponding Z score ("Zmean") from the MERLIN output at each 1 cM position as the linkage "priors" for the weighted association analysis.

2.4.3 Weighted Association Analysis Methods

2.4.3.1 Linkage Weights

The theoretical basis for using linkage weights in the context of genome-wide association analysis comes from the general literature of weighted hypothesis testing (Roeder, Devlin, & Wasserman, 2007). Roeder et al. (2006) proposed using weights from a linkage scan to attenuate the vast multiple testing pitfalls encountered with GWAS data. Investigators typically favor one chromosome region or another based on prior evidence – one primary consideration is linkage

data (Roeder et al., 2006). Weighting association statistics using linkage data is a quantitative method of incorporating prior information into large-scale association scans. While there are numerous approaches to devising weights, there are only two criteria that must be satisfied: (1) each weight must be greater than or equal to 0 and (2) the mean of the weights must equal 1. As noted by Roeder et al. (2006), a reasonable weighting choice is to use nonparametric linkage scores generated from linkage scans. Because what constitutes a linkage region in binary terms is often not well defined, it follows that the quantitative linkage signals be used to generate continuous weights. In addition, it is recommended that the GWAS and prior linkage information come from a similar ethnic background, as genetic heterogeneity between ethnic backgrounds will reduce the power gained from using prior information.

In the original description of using linkage statistics to weight association *p*-values, two continuous weighting schemes were introduced - exponential and cumulative. Exponential weights can be highly sensitive to large linkage signals while cumulative weights tend to be less so. Given the large linkage signal in the original combined analysis for the narrow phenotype definition (bipolar I) found on chromosome 6q (LOD=4.2) we chose to use the cumulative weighting scheme in an attempt to weight the association signals in a more evenly distributed manner. Another consideration was the defining the scaling factor, whereby in the context of cumulative weights (as used here) constrains the influence of any individual linkage peak such that the weights will increase linearly for linkage scores (Z) near B, but reach an asymptote for large linkage scores ($|Z - B|$) (Roeder et al., 2006). We chose a scaling factor (B) of 2 based upon recommendations found in Roeder et al. (Roeder et al., 2006). This results in linkage scores greater than or equal to 2 units above B (2) to be approximately equally up-weighted. The same is true for linkage scores less than or equal to 2 units below B (2) in that they are equally down-

weighted.

2.4.3.2 Weighted Association

The weighted association approach incorporates the weighting scheme directly into the association *p*-values. In order to match an association *p*-value with its respective linkage signal, we assigned a genetic distance to each SNP location. As a result, *p*-values for each SNP may be up-weighted or down-weighted depending upon the relative linkage signal in that region. The nominal *p*-values from the association test are divided by the respective weights given to the genomic region (determined by the linkage signal) to generate the weighted *p*-values. The weighted *p*-values are then used in the false-discovery rate (FDR) procedure to assess overall significance. The weighting procedure (and implementation of the FDR adjustment) was conducted in an R (http://www.r-project.org) script entitled, "weighted_FDR.R" (http://www.wpic.pitt.edu/wpiccompgen/fdr). For the FDR method, we used the method described in Storey (2002) with a significance level of 0.10.

2.5 Results

2.5.1 Primary Association Results

A histogram of weights across the genome is shown in figure 2.2. Our weighted association results for the set of 342,191 SNPs along with their respective weights, is shown in figure 2.3 and 2.4. As mentioned in our methods section, we decided to use the cumulative weighting procedure as our official results, but include a graph and weight distribution for the exponential weighting procedure for the sake of comparison. None of the weighted association signals reach genome-wide significance after implementing the FDR method (multiple correction threshold $0.1/342{,}191 = 2.92 \times 10^{-7}$). Our most significant association signal was at rs4939921

on chromosome 18p21.1 (weighted $p = 4 \times 10^{-7}$, weight = 5.2), residing in an intron of myosin VB (MYO5B), and is shown in figure 2.5. While this was also the top signal in the unweighted association, the signal strength increased by almost an order of magnitude through the weighted approach (original $p = 1.9 \times 10^{-6}$). All but one of the neighboring SNPs with weighted $p < .001$ were found to be in linkage disequilibrium (LD) with rs4939921. Our second strongest signal came from rs4852259 on chromosome 2p13.3 (weighted $p = 2.3 \times 10^{-6}$, weight = 3.5), residing in an intron of the gene dysferlin (DYSF), also shown in figure 2.6. This SNP was the third ranked signal in the unweighted association. Unlike the top signal, there were no neighboring SNPs with $p < .001$.

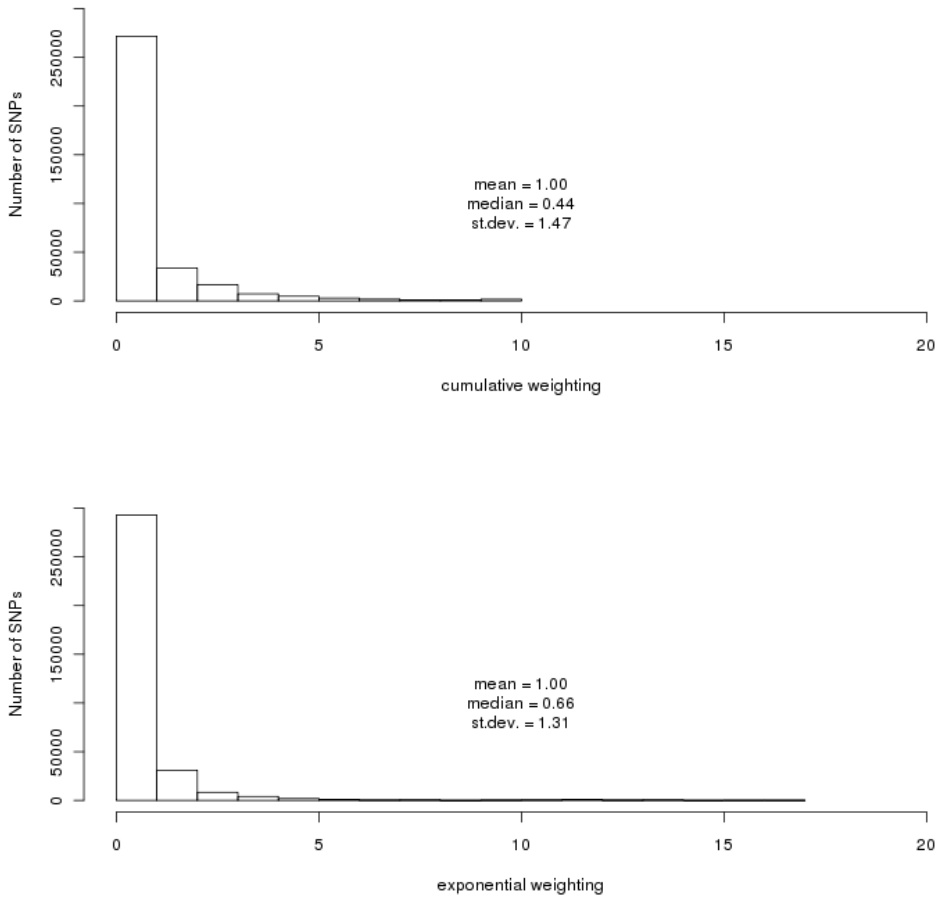

Figure 2.2 Distribution of Weights Across the Genome. Cumulative weighting is shown on top and exponential weighting is shown on bottom. For both weighting schemes, the majority of SNPs (~71%) have a weight less than one, leading them to have a higher (or less significant) *p*-value.

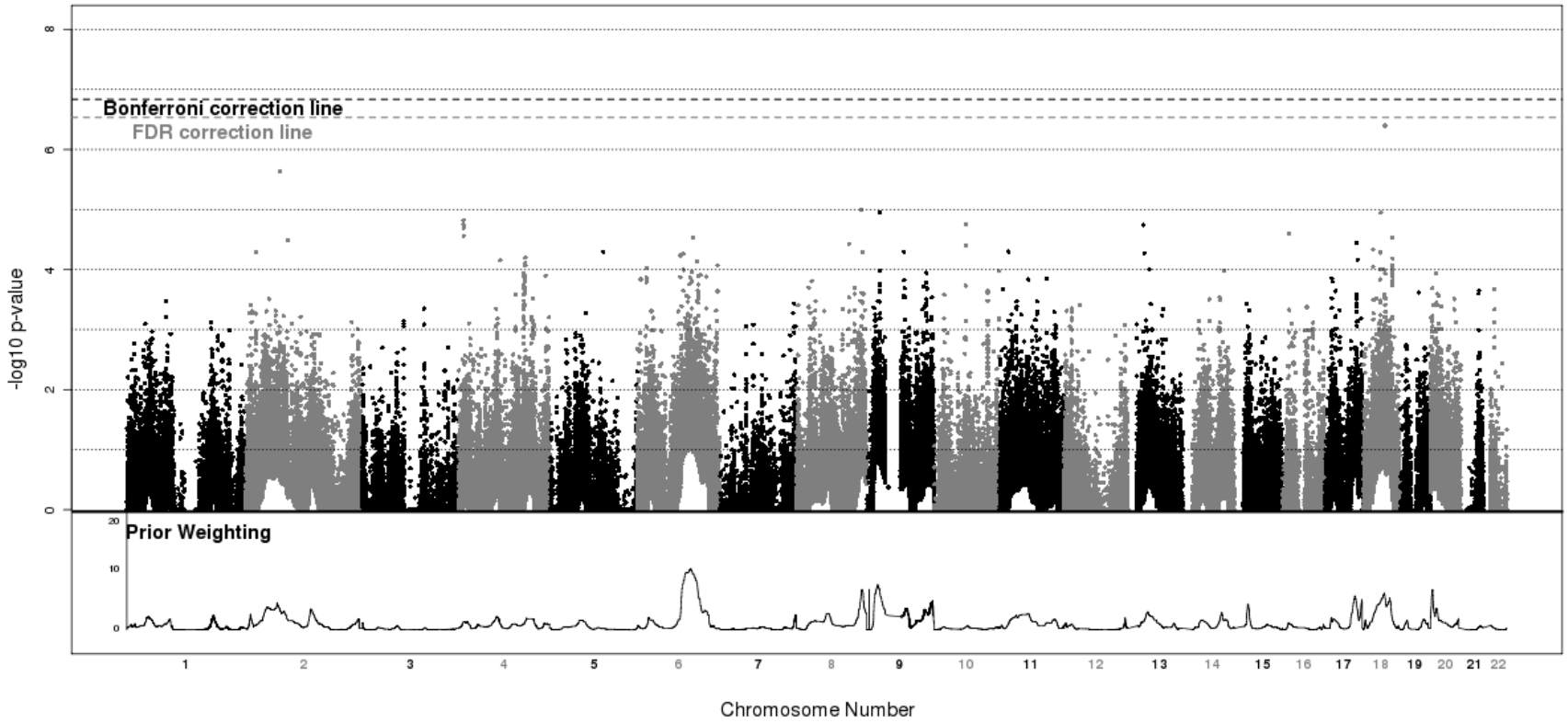

Figure 2.3 Cumulative Weighted Cochran-Armitage Association Results of the STEP-BD Cohort (342,191 SNPs). Cumulative prior weighting shown at bottom of the plot.

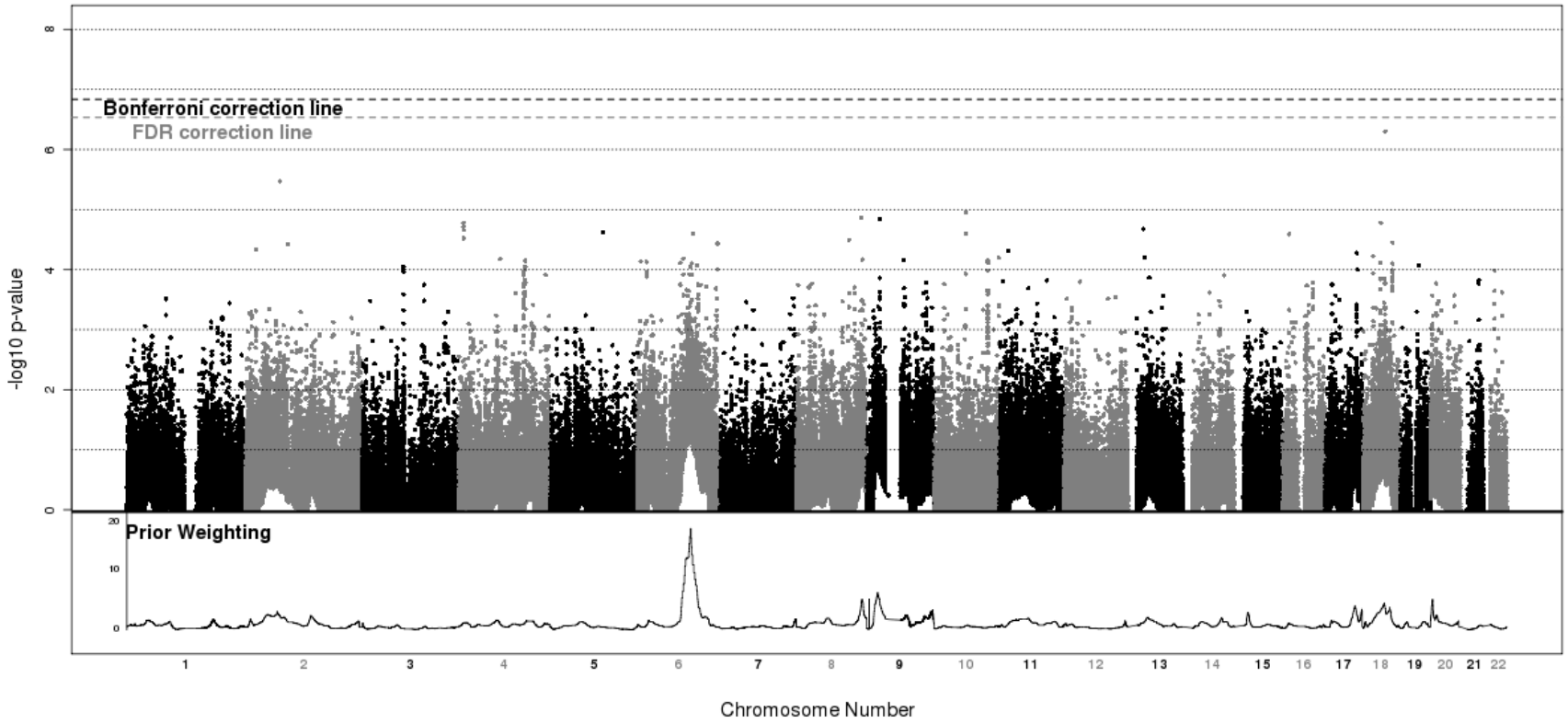

Figure 2.4 Exponential Weighted Cochran-Armitage Association Results of the STEP-BD Cohort (342,191 SNPs). Prior weighting shown at bottom of the plot.

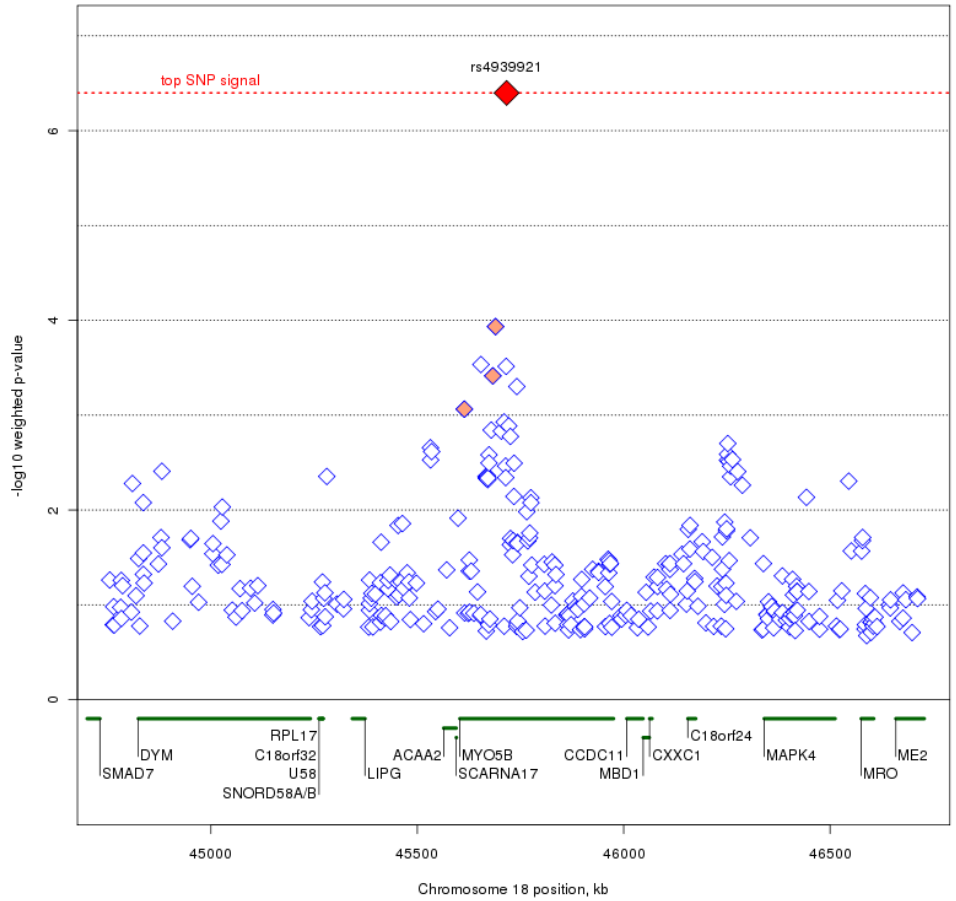

Figure 2.5 MYO5B Signal on Chromosome 18q21.1-18q21.2. The most associated SNP is marked in red. The color of the remaining SNPs reflects linkage disequilibrium (r^2) with the most associated SNP (increasing red hue marks stronger r^2). Gene regions are shown in green, and were taken from the March 2006 UCSC genome browser assembly.

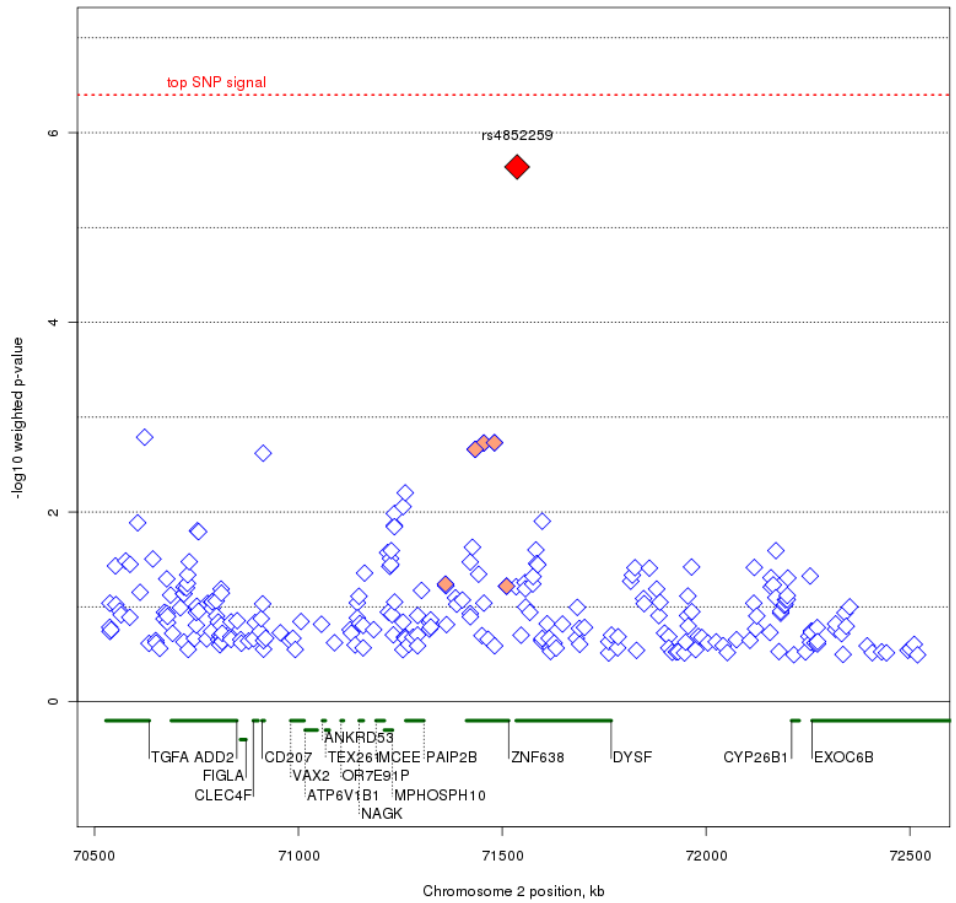

Figure 2.6 DYSF Signal on Chromosome 2p13.3. The most associated SNP is marked in red. The color of the remaining SNPs reflects linkage disequilibrium (r^2) with the most associated SNP (increasing red hue marks stronger r^2). Gene regions are shown in green, and were taken from the March 2006 UCSC genome browser assembly.

2.5.2 Up-weighted Regions of Interest

Beyond our top SNP signals, we also looked for regions where linkage signals coincided with clusters of top SNP signals. Specifically, we focused on SNP signals with weighted $p < .001$ that resided in the top 10th percentile of weight values across the genome (figure 2.7). The majority of SNPs that fit these criteria occur on Chromosomes 2, 6, 8, 9, 17, 18, and 20. Previous work using the STEP-BD has been done on the up-weighted region seen on chromosome 6 (Fan et al. 2010), so our focus here is on the remaining chromosomes. In particular, we looked for clusters of neighboring SNPs showing a consistent signal. We've highlighted three regions of interest (figures 2.8 - 2.10) on chromosome 9p21.2, 17q24.2, and 18q12.2. The strongest signal on the chromosome 9p21.2 region, rs17760820, resides in an intron of open reading frame C9orf82, with four nearby SNPs at $p < .001$ residing in introns of the intraflagellar transport 74 homolog (IFT74). The two genes are 54 kb apart, and the SNPs in IFT74 are in moderate LD with the SNPs in C9orf82. On chromosome 17q24.2, the notable cluster of neighboring signals all reside in an intergenic region. The top signal, rs16974356, is 220 kb telomeric of the mitogen-activated protein kinase kinase 6 (MAP2K6) gene, and 312 kb centromeric of the potassium inwardly-rectifying channel, subfamily J, member 16 (KCNJ16) gene. All the SNPs in this region with weighted $p < .001$ are in relatively high LD in this intergenic region. The SNP cluster on chromosome 18q12.2 also resides outside of a gene, with the nearest transcript being hypothetical protein LOC647946 where the 5' end is 199 kb away from the top signal in the region, rs2862294. Although most of the nearby SNPs with weighted $p < .001$ are in LD with rs2862294, a few SNPs show no LD despite their close proximity.

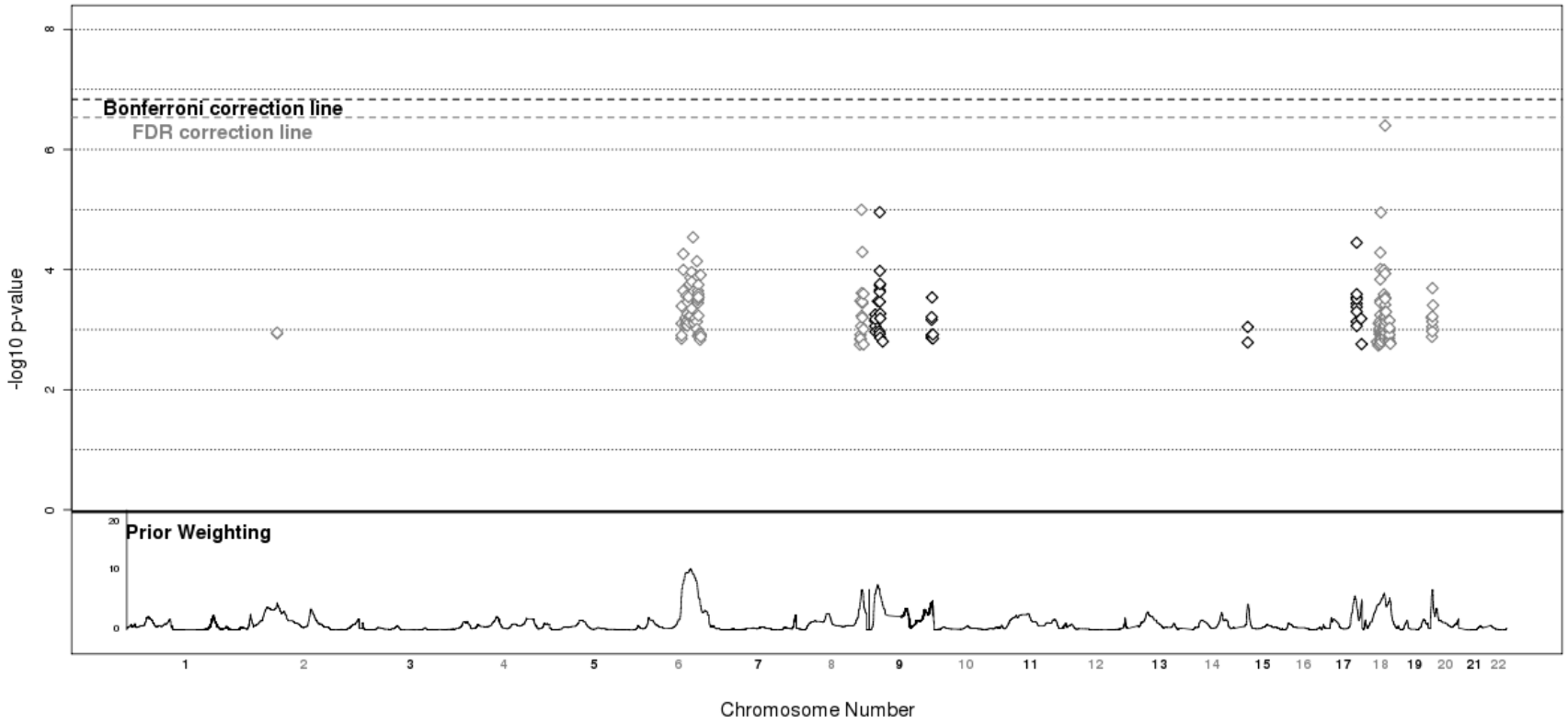

Figure 2.7 Association Plot of all SNPs with Weighted $p < .001$ and in the Top 10th Percentile of Linkage Weights. Prior weighting shown at the bottom of the plot.

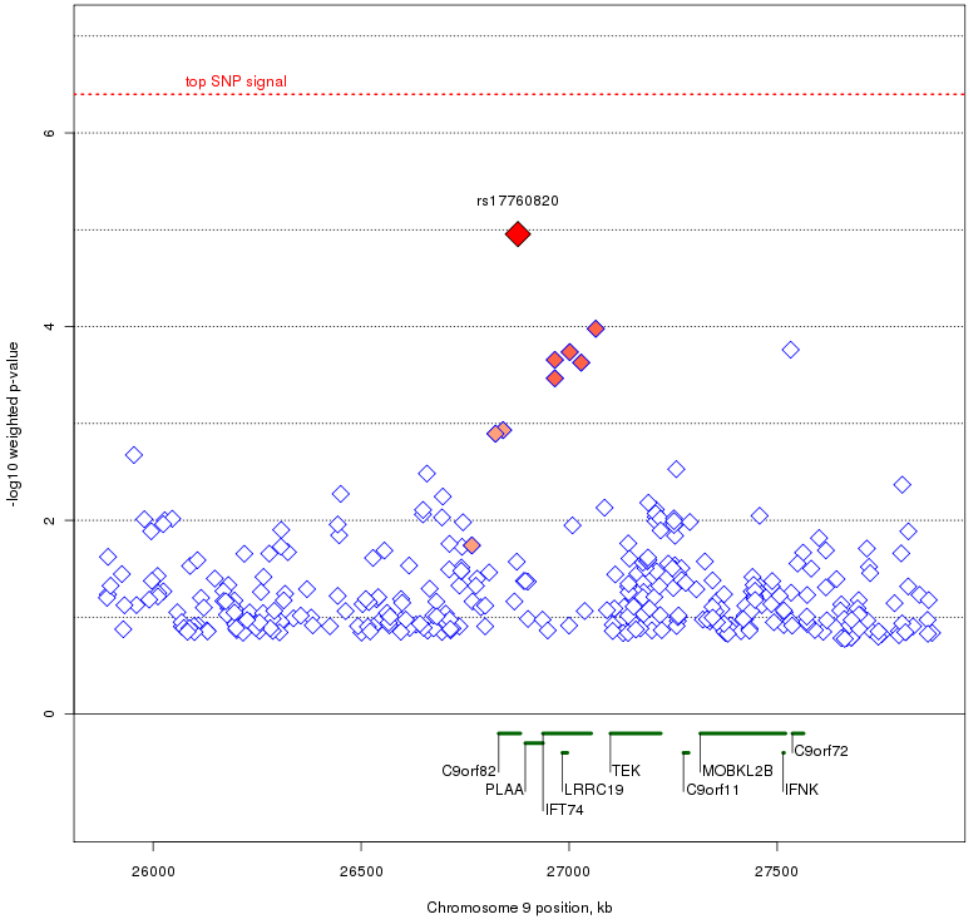

Figure 2.8 Chromosome 9p21.2 Region. The most associated SNP is marked in red. The color of the remaining SNPs reflects the linkage disequilibrium (r^2) with the most associated SNP (increasing red hue marks stronger r^2). Gene regions are shown in green, and were taken from the March 2006 UCSC genome browser assembly.

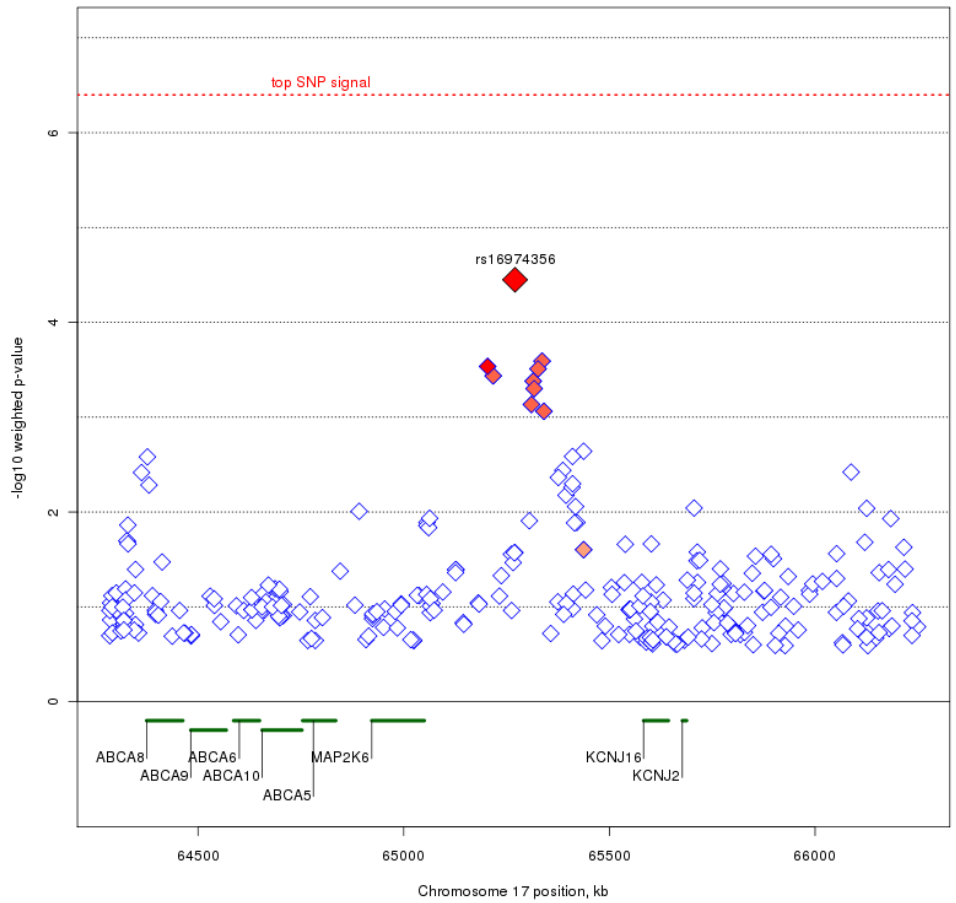

Figure 2.9 Chromosome 17q24.2 Region. The most associated SNP is marked in red. The color of the remaining SNPs reflects the linkage disequilibrium (r^2) with the most associated SNP (increasing red hue marks stronger r^2). Gene regions are shown in green, and were taken from the March 2006 UCSC genome browser assembly.

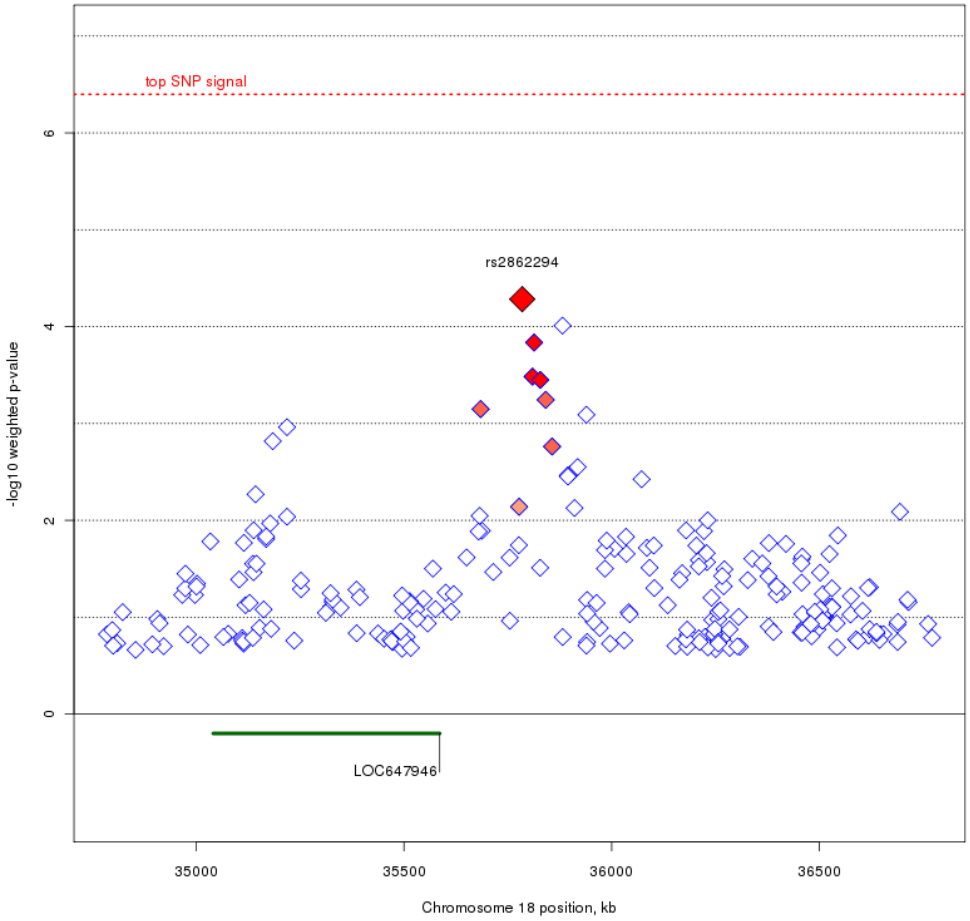

Figure 2.10 Chromosome 18q12.2 Region. The most associated SNP is marked in red. The color of the remaining SNPs reflects the linkage disequilibrium (r^2) with the most associated SNP (increasing red hue marks stronger r^2). Gene regions are shown in green, and were taken from the March 2006 UCSC genome browser assembly.

2.6 Discussion

The primary goal of this study was to examine how prior linkage information could inform association signals across the genome. Using a linkage-weighted FDR approach, we've highlighted particular regions that would not have been addressed by association alone. Although no signal reached the conservative criteria of genome-wide significance, we report on a few areas where linkage evidence and strong association signals coincide.

Our strongest association signal, rs4939921 in MYO5B, was expected, as it was also the top SNP prior to the weighted association. In addition, this was the top signal reported in Sklar et al. (2008) where the STEP-BD sample was combined with another large BPD sample from the University College of London. MYO5B is a brain-expressed gene that is involved in protein transport and vesicle trafficking at the plasma membrane (Lapierre & Goldenring, 2005; Lise et al., 2006), RNA transcription (Lindsay & McCaffrey, 2009), and has recently been implicated in microvillus inclusion disease, a rare genetic disorder of the small intestine (Erickson, Larson-Thome, Valenzuela, Whitaker, & Shub, 2008; Muller et al., 2008). Although our prior linkage information up-weights this signal at MYO5B, a recent meta-analysis combining the STEP-BD sample with a number of other case-control BPD samples finds no evidence of association in this region (Ferreira et al., 2008). Our second strongest signal, rs4852259 in DYSF, resides in a gene that codes for skeletal muscle protein, and non-synonymous mutations in DYSF have resulted in limb girdle muscular dystrophy and Miyoshi myopathy (Bashir et al., 1998; Liu et al., 1998). In our literature review of DYSF, we did not find any evidence of brain expression or psychiatric effects of this gene.

We also identified genes in regions where prior linkage information and association signals coincide. At chromosome 9p21.1, signals reside in both C9orf82 and IFT74. While there

is currently no known information on C9orf82, IFT74 (also known as capillary morphogenesis protein 1: CMG1) is a brain-expressed gene, whereby it codes a protein that transports material from the cell body along the dendritic and axonal processes of neurons (Momeni et al., 2006). IFT74 has been implicated in familial cases of amyotrophic lateral sclerosis (ALS) - frontotemporal dementia (FTD), although common SNPs do not seem to be implicated as causal variants of ALS-FTD (Xiao et al., 2008). On chromosome 17q24.2, none of the association signals occur within or adjacent to genes (all signals with weighted $p < .001$ are greater than 100 kb away from the nearest gene). The nearest genes that flank the signal are MAP2K6, part of the kinase mediated signal transduction pathway involved in cell cycle arrest, transcription activation, and apoptosis; and KCNJ16, coding a membrane protein that regulates potassium channel activity. Of note, MAP2K6 has been associated in mediating the onset of Huntington's disease (Arning et al., 2008), and KCNJ16 reflects ion channel activity, which has recently been implicated as a possible source of BPD pathogenesis (Ferreira et al., 2008). We can only speculate that the association signal may be involved in some sort of regulatory function relative to these genes. For the signal on chromosome 18q12.2, the current UCSC genome browser contains only a hypothetical gene, location 647946, in the region. Further annotation is necessary to determine whether location 647946 is indeed a coding region, and no functional information is currently known about it.

2.6.1 Conclusion

By incorporating genome-wide linkage and association data on BPD, we hope to identify susceptibility alleles and genes that would have otherwise been overlooked by any single method of genome-wide interrogation. Due to the conservative nature of genome-wide significance in association designs, true signals of small effect are inevitably overlooked. Prior linkage data

serves to aid in their detection as it provides independent marker information on BPD across the genome. We should note that while our prior linkage information came from a substantial pool of linkage studies, this method could certainly be applied with data from a single linkage study. Although weighted designs with informative prior information maximize any potential power gains, the loss in power is small when using uninformative prior information (Roeder et al., 2006). Given that linkage signals are better suited to detect rarer alleles of high penetrance, we believe this method will be even more helpful as future GWAS increase the coverage of rarer SNP and CNV variants (MAF < .05) to their analyses. Furthermore, recent work using simulation and known genetic associations postulates that signals found in GWAS could actually be "synthetic associations" driven by rare genetic variants (Dickson, Wang, Krantz, Hakonarson, & Goldstein, 2010). In this respect, combining linkage and association data can be useful in highlighting areas that are most promising for future sequencing studies; studies that would be able to putatively identify the effects of rare variants on BPD.

ACKNOWLEDGEMENTS
DP Howrigan's contribution to this work was partially supported by two institutional training grants from the National Institute of Child Health and Human Development (T32 HD007289, Michael C. Stallings, Director) and the National Institute of Mental Health (T32 MH016880, John K. Hewitt, Director) awarded to the Institute for Behavioral Genetics, University of Colorado. JW Smoller's contribution to this work was supported in part by the National Institute of Mental Health (MH063445).

Chapter 3

Detecting Autozygosity through Runs of Homozygosity: A Comparison of Three Autozygosity Detection Algorithms

3.1 Background

With the advent of high-density genome-wide SNP arrays, examination of individual genetic data has revealed that runs of homozygosity (ROHs) - many homozygous SNPs in a row - are a common occurrence in all populations worldwide (Gibson, Morton, & Collins, 2006). Consequently, there has been interest in understanding if ROHs serve as risk factors underlying complex and simple disorders. There are sound theoretical reasons to suspect that ROHs are associated with disorder risk. Long ROHs (e.g., 100+ homozygous SNPs in a row) are unlikely to have arisen by chance. Rather, they are likely to denote autozygosity, which occurs when two genetic strands in the same individual come from the same ancestor - in other words, when (perhaps distant and unintended) inbreeding occurs. Inbreeding has long been known to increase the risk of many disorders.

Much research suggests that such "inbreeding depression" occurs via an increase in autozygosity and a corresponding increase in homozygosity at rare, partially recessive, deleterious mutations (reviewed in Charlesworth & Willis, 2009)). In order for researchers to investigate the effects of autozygosity on disease, it is critical to accurately distinguish truly autozygous ROHs from the larger pool of often non-autozygous ROHs in a sample. The goal of this study is to investigate how accurately existing ROH detection programs identify autozygosity in genome-wide SNP data and which thresholds within these programs maximize the ability to detect genomic signatures of inbreeding depression.

3.1.1 Review of ROH Literature

ROH analyses to date have investigated questions relevant to both basic population genetic theory and disease risk. Population genetics studies have analyzed the distribution, prevalence, and location of ROHs across various sub-populations to infer population structure, history, and natural selection (Gibson et al., 2006; Kirin et al., 2010; Li et al., 2006; McQuillan et al., 2008; Nothnagel, Lu, Kayser, & Krawczak, 2010; Sabeti et al., 2007; Voight, Kudaravalli, Wen, & Pritchard, 2006). Phenotypic studies have used both family-based and population-based samples to identify specific associated risk ROHs as well as differences in overall ROH burden. There has been recent success in identifying genes underlying simple autosomal recessive disorders in families within populations with high consanguinity using homozygosity mapping, with dozens of publications in recent years (e.g., Abu Safieh et al., 2010; Borck et al., 2011; Collin et al., 2010; Kalay et al., 2011; Walsh et al., 2010; Zelinger et al., 2011). Larger scale studies using genome-wide SNP data have also been conducted for complex phenotypes such as Schizophrenia (Lencz et al., 2007), Bipolar Disorder (Vine et al., 2009), Parkinson's disease (Wang, Haynes, Barany, & Ott, 2009), Alzheimer's disease (Nalls et al., 2009), Colorectal cancer (Spain, Cazier, Houlston, Carvajal-Carmona, & Tomlinson, 2009), Childhood acute lymphoblastic leukemia (Hosking et al., 2010), and Breast and Prostate cancer (Enciso-Mora, Hosking, & Houlston, 2010). Non-clinical traits such as height (Yang et al., 2010b) have also been examined using ROH analyses (for a review of the current ROH research, see Ku, Naidoo, Teo, & Pawitan, 2010). However, results from these previous studies on complex phenotypes have been mixed. While significant ROH have been identified for height and Alzheimer's disease, little to no evidence exists for the effects of specific ROH on other phenotypes. Moreover, the effects of ROH burden on some complex phenotypes (Schizophrenia, Alzheimer's

disease) were significant, whereas no effects of ROH burden were found on other complex phenotypes (Bipolar Disorder, Colorectal cancer, Childhood acute lymphoblastic leukemia, Breast cancer, and Prostate cancer).

3.1.2 Limitations of ROH Analysis

A central limitation to current studies analyzing ROHs is the lack of consensus criteria or even guidelines for defining a ROH (Ku et al., 2010). For example, Lencz et al. (2007) only examined ROHs shared by ten or more subjects and that spanned at least 100 SNPs, and did not allow for any heterozygote calls, whereas Spain et al. (2009) examined overall ROH burden across various SNP and kb length thresholds, analyzed both complete and low linkage disequilibrium (LD) SNP datasets, and permitted a 2% heterozygote allowance. The discrepancy between definitions of ROHs makes comparisons between study results difficult, and the lack of consensus criteria for defining ROH increases the probability of false positive results due to the potential for choosing the most significant among many ROH thresholds investigated (Ioannidis, 2005).

3.1.3 ROH as an Optimal Measure of Autozygosity

A recent study by Keller, Visscher, & Goddard (2011) found that inbreeding coefficients estimated from ROHs are much better at detecting the overall burden of rare, recessive mutations (the likely cause of inbreeding depression; Charlesworth & Willis, 2009) than several alternatives, including inbreeding coefficients defined on a SNP-by-SNP basis and those defined from pedigrees. While SNP-by-SNP homozygosity provides an adequate test for recessive effects of common causal alleles (that either exist on the genotyping platform or that are in LD with genotyped SNPs), ROHs track autozygosity, and therefore can be used to investigate the effects of homozygosity at both rare and common causal variants (Keller et al., 2011). Non-

autozygous ROHs, stretches of homozygous SNPs that are actually heterozygous at unmeasured variants, are less likely to contain rare, partially recessive, deleterious mutations in their homozygous form. Therefore, the central criterion for defining ROHs – and the only reason one would measure ROHs rather than SNP-by-SNP homozygosity – is to assess autozygosity. In practice, this means differentiating ROHs that are not autozygous and are identical-by-state (IBS) from ROHs that are autozygous and are identical-by-descent (IBD). However, there has been no systematic investigation to date into which ROH detection program is optimal at detecting autozygosity and which parameters within those programs maximize statistical power. The current study addresses these unanswered questions and offers some consensus criteria to capture autozygosity through ROH analysis.

3.2 Methods

3.2.1 Overview of Approach

Our analysis simulated sequence data that mimicked LD and polymorphism properties found in modern European heritage populations, thus allowing the sequence to resemble expected autozygosity in an outbred population as well as provide perfect information about truly autozygous segments. SNP data was obtained from the sequence by sampling common polymorphisms that mimicked the allele frequency distribution and SNP density found in a modern dense SNP chip (e.g., the Affymetrix 6.0 SNP chip), adding error rates and missingness patterns that were also empirically derived. Using this SNP data, we evaluated the performance of existing ROH detection programs to detect known autozygous segments. There are three primary ROH detection programs that have either been used in previously cited ROH studies and/or that have been the focus of a recent publication on detection of autozygosity: PLINK

(Purcell et al., 2007), GERMLINE (Gusev et al., 2009), and BEAGLE (Browning & Browning, 2010). To assess how accurately each program identified autozygosity, we estimated the rate at which non-autozygous ROHs were called "autozygous" (type 1 errors) and the rate at which truly autozygous ROHs were not detected (type 2 errors).

While low type 1 and type 2 error rates are always preferred, they cannot be minimized simultaneously: an inherent trade-off exists such that an increase in the type 1 error rate leads to a decrease in the type 2 error rate and vice versa. Determining which ratio of type 1 to type 2 errors should be preferred is not obvious; here, we used a second, independent simulation to find which ratio of type 1 and type 2 error rates would maximize power to detect an association between autozygosity burden and a simulated phenotype. We started by simulating a phenotype associated with autozygosity, and from this population drew a sample containing autozygous segments at the rate found in our simulated sequence data (i.e., the level of autozygosity that corresponds to ROH distributions seen in empirical data). We then regressed the simulated phenotype on the sum of segments identified as "autozygous," which included truly autozygous segments (influenced by the type 2 error rate) as well as non-autozygous segments (type I errors), as indicated by the type 1 and type 2 error rates found for the program/thresholds from the previous analysis. Power in this case is defined as the proportion of significant results observed in the simulation. By comparing power across programs and across thresholds within those programs, we have an empirical, objective foundation for deciding which program and which thresholds are most suitable for detecting autozygous ROHs.

3.2.2 Generation of Sequence Data

In order to test the performance of detecting autozygous segments for each program, we needed genomic data that identified which segments of some arbitrary length were truly

autozygous from a common ancestor within some time frame (e.g., 50 generations). One method would be to use real sequence data in existing samples, but detecting autozygosity in sequence data given current small samples, low pass coverage, and high error rate estimates (e.g., 1-3% in the thousand genomes data; Durbin et al., 2010) poses a substantial problem in accurately estimating autozygosity. Instead, we generated sequence data (genomic data with every base pair measured) that tracked every allele, rare or common, in the population, allowing us to identify autozygous segments without error by finding genomic areas of some arbitrary length (e.g., 0.5 Mb or larger) that were perfectly (100%) homozygous at the sequencing level.

We used the forward-time simulation program FREGENE (Chadeau-Hyam et al., 2008) to simulate full sequence data. FREGENE simulates a monoecious diploid population that evolves over non-overlapping generations according to a Wright-Fisher model (Fisher, 1930). We simulated a 120 Mb chromosome in an effective population (Ne) of constant size 10,000 (roughly the estimated effective population size of humans; Harpending et al., 1998) for 100,000 generations, long enough for mutation-drift equilibrium to be assured (Chadeau-Hyam et al., 2008). Using neutral simulation parameters recommended by FREGENE, we set the mutation rate at 2.3e-8, the gene conversion rate at 4.5e-9 with a 500 bp gene conversion length, and no variants under selection. The average recombination rate was 1.3e-8, but FREGENE allows for realistic differences across the genome in recombination, with most (80%) recombination occurring in hotspots of length ~2,000 bp that encompassed 20% of the chromosome (Chadeau-Hyam et al., 2008; Schaffner et al., 2005). It was critical to simulate patterns of LD that mimicked as closely as possible those observed in real human SNP data, as the lengths, distributions, and frequencies of truly autozygous ROHs and non-autozygous ROHs are influenced by the population size, the degree of real inbreeding in the population, and the LD

patterns between SNPs. For example, an isolated population with long haplotypes and long distance LD would exhibit a high proportion of long ROH even if few arose from recent inbreeding. For both the sequence and resulting SNP data parameters, we used SNP data from control subjects in the Molecular Genetics of Schizophrenia nonGAIN sample (Shi et al., 2009) as our empirical SNP data set to check the validity of the simulation. The empirical data was ascertained on an Affymetrix 6.0 SNP chip and contained ~770,000 SNPs across the ~2,770 Mb of the autosomal portion of the human genome that is 'SNP-mappable.'

LD in data simulated under a neutral mutation-drift model is known to have much lower LD than is observed in human data (Reich et al., 2001; Schaffner et al., 2005). Both Reich et al. (2001) and Schaffner et al. (2005) found that one or two population bottlenecks between 800 to 3,000 generations ago led to LD patterns that mimic data from a population of European heritage. Older bottlenecks led to less LD and more recent ones to more LD. Starting from the population of Ne=10,000 in mutation drift-equilibrium, we varied bottleneck parameters until we found those that mimicked real LD patterns when we sampled SNP data from the sequence. We found that a population bottleneck reduction from 10,000 to 800 individuals for 200 generations, followed by 2,000 generations of evolution back at 10,000 best mimicked the LD patterns seen in our empirical SNP data (see Figure 3.1), and similar to the results seen in Reich et al. (2001).

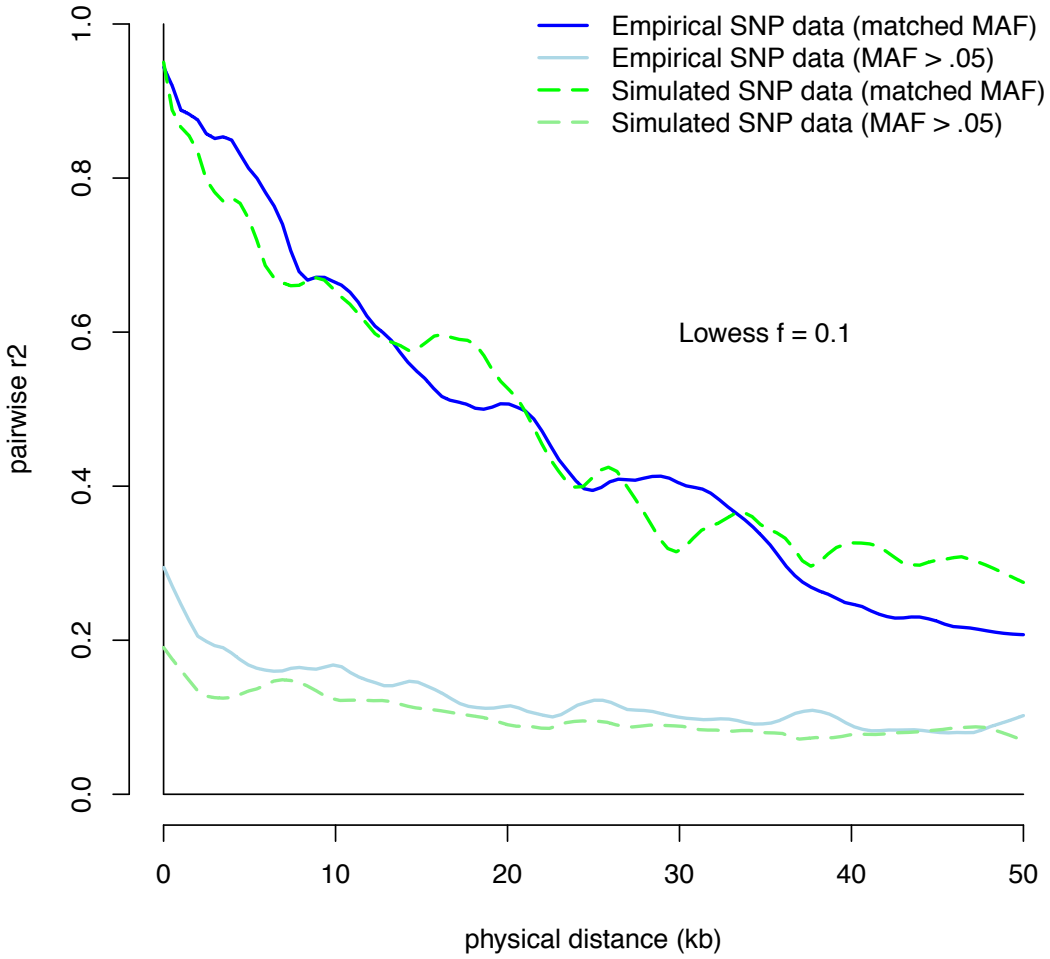

Figure 3.1 Observed LD within 50 kb using 5,000 Pairwise SNP Comparisons. LD patterns are measured by sampling the r^2 of 5,000 SNP pairs up to 50 kb away. The lines represent Lowess curves that track changes in mean r^2 as physical distance increases between SNPs. Empirical r^2 values are in blue, and simulated r^2 values are in green. Because MAF strongly influences the r^2 between SNPs, a floor effect often occurs after short distances due to SNP pairs with large MAF differences, making it difficult to compare LD patterns between datasets. Therefore, we included both closely matched MAF SNP pairs and SNP pairs with minimum MAF > .05. Darker hues represent matched MAF pairs and lighter hues represent SNP pairs with minimum MAF > .05.

We then turned our attention to generating simulated data that led to similar lengths and frequency of ROHs as seen in our empirical SNP data. We found that reducing population size to 6,500 (from 10,000) and selective sampling of individuals from that reduced population best mimicked the length of ROHs seen in our real data. In particular, we chose 1,000 individuals from the sequence data that closely matched the overall proportion of ROH seen in our empirical data across various ROH analyses. Thus, we simulated genetic sequence data that mimics as closely as possible the two parameters – LD and distribution of ROHs – central to the present investigation. Our sample of simulated sequence data contained 669,219 total variants, with 436,564 having a minor allele frequency (MAF) > 1%, and on average one variant per 179 bp.

3.2.3 Mapping of Autozygous Segments

It would be ideal to keep track of autozygous segments through the course of our sequence simulation. Unfortunately, no sequence simulation program that we are aware of tracks autozygosity. As a substitute that will detect all but the shortest autozygous tracks, we identified autozygous segments by finding stretches of sequence data that were perfectly homozygous. To do this, we first used the genetic distance map derived from the FREGENE simulation to estimate the expected length of autozygous segments. By definition, both genetic strands making up an autozygous segment originate from a single common ancestor, however the length of this segment decreases on average over time due to recombination. Specifically, the expected length of an autozygous segment follows an exponential distribution with mean equal to $1/2g$ Morgans, where g is the number of generations since the common ancestor. Thus, the expected length of an autozygous segment caused by sib-sib inbreeding ($g = 2$, counting from the inbred offspring to the siblings who mate, and the mated siblings to their parents) is 1/4 Morgan or 25 cM, while the expected length of an autozygous segment originating from a common ancestor 50 generations in

the past is 1 cM (see Figure 3.2).

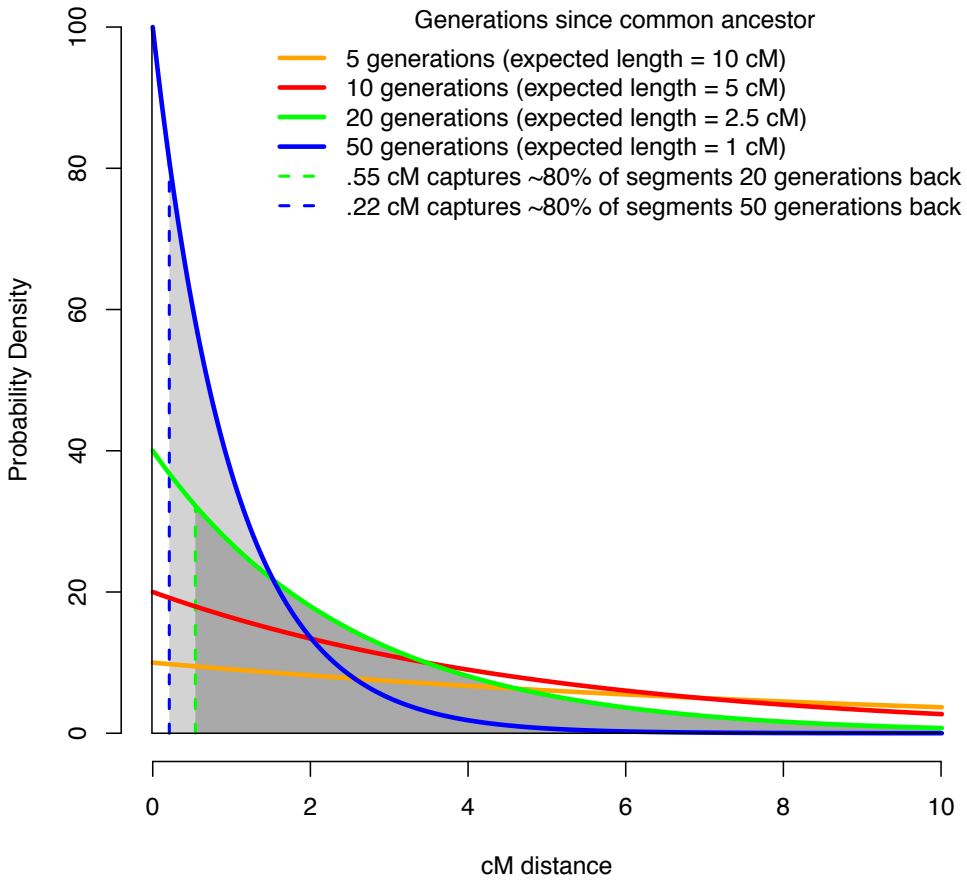

Figure 3.2 Distribution of Expected Autozygous Segment Lengths since Common Ancestor. The probability density of autozygosity lengths from a common ancestor follows an exponential distribution that depends on the number of generations since the common ancestor. We chose to examine autozygous segments that originate from a common ancestor within the past 20 generations (in green) and within the last 50 generations (in blue). The dashed lines indicate the minimum length threshold that would capture 80% of these segments, with the grey area under the curve being the proportion of autozygous segments being captured above this threshold.

Because the shortening of autozygous lengths due to recombination occurs gradually across generations, any choice of distance threshold to define autozygosity is ultimately arbitrary. We chose two thresholds that were consonant with the lengths of homozygous runs being reported in the literature and that were realistically detectable using modern SNP platforms. The first (long) threshold captured 80% of autozygosity originating in common ancestors within the past 20 generations (~600 years in humans), and the second (short) threshold captured 80% of autozygosity originating within the past 50 generations (~1500 years ; Fenner, 2005). As shown in the areas under the curve in Figure 3.2, these thresholds correspond to a minimum genetic distance of 0.55 cM (~423 kb) for 20 generations back and 0.22 cM (~169 kb) for 50 generations back, where the genetic distance derived from FREGENE recombination is 1.3 cM/Mb. By requiring segments in our sequence data to be completely homozygous, new mutations arising within the last 20 or 50 generations on either segment would cause regions to be missed that were truly autozygous. To ensure that autozygous segments were completely homozygous and therefore detectable with 100% fidelity, we allowed no new mutations to arise during the final 50 generations of the simulation. Such a 'mutational freeze' has a negligible impact on the resulting SNP data, as recent neutral mutations very rarely rise in frequency to be considered SNPs (MAF > 1%; Hedrick, 2011). On the other hand, the mutational freeze did affect sequence data, but given that sequence data was only used for inferring autozygosity, this strategy did not affect our conclusions. Within the past 20 generations, the average autozygous segment spanned 4,707 variants and 841 kb in length, with total autozygosity covering 0.36% of the sequence data. Within the past 50 generations, the average autozygous segment spanned 1,862 variants and 334 kb in length, with total autozygosity covering 0.91% of the sequence data.

3.2.4 Extracting SNP Data from Sequence Data

We extracted a subset of variants from the simulated sequence data to mimic several properties found in empirical SNP data measured on a modern, commonly used SNP platform (the Affymetrix 6.0 SNP chip). We first sampled a subset of SNPs that matched the MAF distribution from our empirical data (see figure 3.3). We then extracted SNPs that matched the spatial SNP density found in a 120 Mb portion of chromosome 5 (which was typical of other genomic regions) of our empirical SNP data, giving us 33,040 SNPs (See figure 3.4). We then added 'genotyping errors' through random sampling of the SNP data at a low (0.2%) and high (1%) rate, the low rate corresponding to error rates observed in Rabbee & Speed (2006) and the high rate corresponding to the error rate of a small number of duplicate genotyped individuals in our empirical data (data not shown). We used error probabilities informed by the discordant calling rates observed by Rabbee & Speed (2006), with heterozygous SNPs being called homozygous at a roughly threefold higher rate than homozygous SNPs being called heterozygous. Missingness was then added through random sampling by converting SNPs to missing values based on missingness rates (0.8%) seen in the empirical data. Finally, we applied standard GWAS cleaning procedures (dropping individuals with SNP missing rate > 5%, dropping SNPs with missingness rate > 2%, dropping SNPs with MAF < 1%, and dropping SNPs out of HWE where chi-square $p < 0.0001$; Sullivan & Purcell, 2008), resulting in a dataset of 30,113 SNPs in the low error rate data (2,927 SNPs removed), and 30,110 SNPs in the high error rate data (2,930 SNPs removed). PLINK, GERMLINE, and BEAGLE used these two SNP datasets for their ROH analyses.

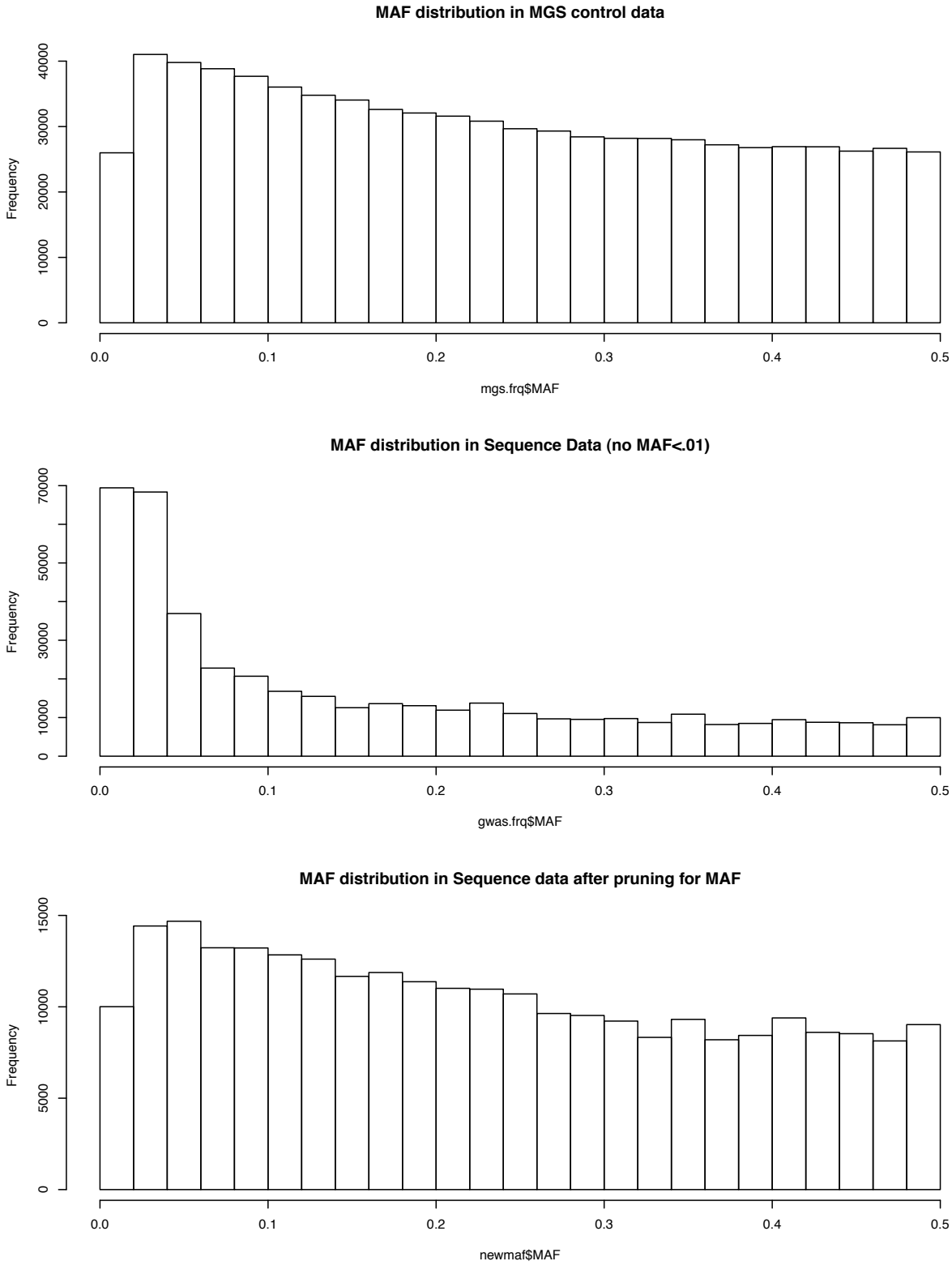

Figure 3.3 MAF Distribution of Sequence and SNP Data Simulation

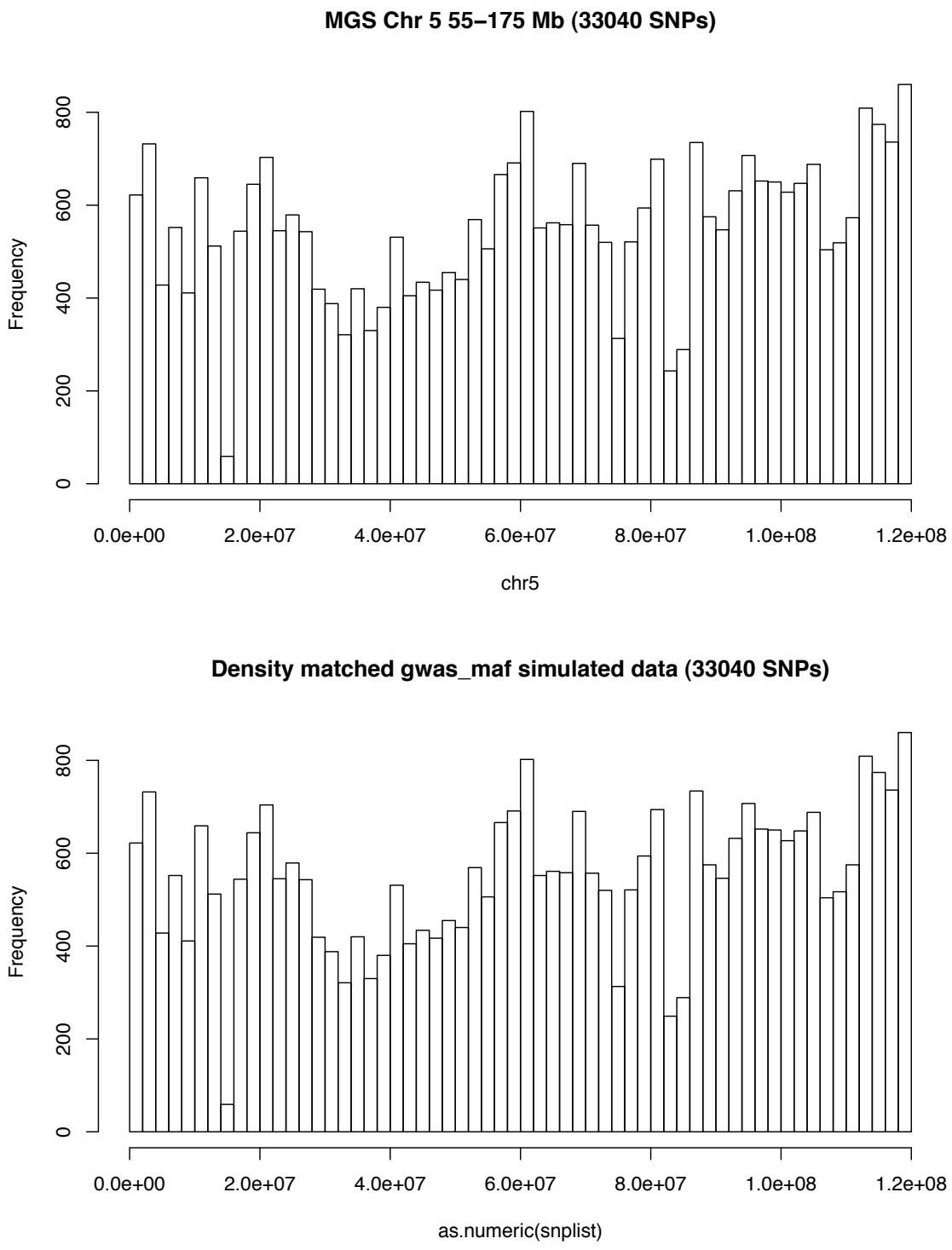

Figure 3.4 SNP Density Matching. After MAF pruning of the simulated sequence data, SNPs were drawn to closely match SNP base positions observed in empirical SNP data

3.2.5 ROH Detection Algorithms

3.2.5.1 PLINK

The –homozyg option in PLINK v1.07 (Purcell et al., 2007) makes ROH calls using a sliding window that scans along an individual's SNP data to detect homozygous stretches. PLINK first determines whether a given SNP is potentially in a ROH. To call a SNP as part of a ROH, PLINK calculates the proportion of completely homozygous windows that encompass that SNP. For example, a SNP inside a 100 SNP window has 100 chances to be part of a homozygous stretch as the window slides across one SNP at a time. Using the default –homozyg-window-threshold of 0.05, if 5% of these windows are completely homozygous, then the SNP will be included in the ROH. Finally, a ROH is called if the number of such "ROH SNPs" in a row surpasses a user-defined threshold in terms of SNPs (default=100) and/or kb distance (default=1,000). PLINK provides numerous other user-defined parameters, such as the size of the sliding window measured in units of SNP length (default=50), the number of heterozygous SNPs (default=1) allowed in the ROH, the number of missing SNPs (default=5) allowed in the ROH, and several other parameters detailed on the PLINK website (see http://pngu.mgh.harvard.edu/~purcell/plink/ and Table 3.1). Of note, the --homozyg-window-kb option in PLINK, which defines windows in terms of distance rather than SNPs, is currently non-functional.

Table 3.1 ROH Detection Parameters. Each program can be run using a command line prompt allowing for each tuning parameter to be defined

Program	Parameters	Code	Parameters Used
PLINK	Varied Parameters:		
	- Heterozygote allowance	--homozyg-window-het	0/1
	- SNP threshold to call a ROH	--homozyg-snp	15/25/50/75/100/200/350/500
	- Sliding window size in SNPs	--homozyg-window-snp	Same as SNP threshold
	- Missing SNP allowance	--homozyg-window-missing	(5% of SNP threshold)
	- Window threshold to call a ROH	--homozyg-window-threshold	(0.05% of SNP threshold)
	Fixed Parameters:		
	- Sliding window size in kb	--homozyg-window-kb	0 (unused)
	- Kb threshold to call a ROH	--homozyg-kb	0 (unused)
	- Minimum SNP density to call a ROH	--homozyg-density (kb)	5,000 (set high to ignore)
	- Maximum gap before splitting ROH	--homozyg-gap (kb)	5,000 (set high to ignore)
GERMLINE	Varied Parameters:		
	- Mismatching heterzygote allowance	-err_het	0/1
	- Minimum ROH length (in cM or Mb)	-min_m (in cM)	0.15/0.25/0.5/0.75/1/2/3.5/5
	- Window size in SNPs	-bits	Expected number of SNPs for given cM length
	Fixed Parameters:		0
	- Mismatching homozygote allowance	-err_hom	
BEAGLE	Varied Parameters:		
	- Non-HBD to HBD transition rate	nonhbd2hbd=	0.0001/0.01/0.1
	- HBD to non-HBD transition rate	hbd2nonhbd=	0.25/0.5/1
	Fixed Parameters: None		

3.2.5.2 LD-Pruning

PLINK does not account for MAF or LD in its algorithm. Aside from the ROH tuning parameters available in PLINK, taking into account MAF and LD in SNP data will also affect how ROH are identified. In particular, many low MAF SNPs in a row can increase the probability of chance (non-autozygous) ROH segments, and high LD within dense SNP regions can also have this effect. To minimize the probabilities of spurious ROH calls, we used LD-pruned data (as suggested in the PLINK manual), such that we first removed SNPs with MAF < 0.05, and then used PLINK's --indep command to prune for LD at two levels, which we term "moderate" and "heavy" LD pruning. Moderate LD pruning removed SNPs within a 50 SNP window that had $r^2 > 0.5$ (corresponding to a variance inflation factor, VIF, greater than 2) with all other SNPs in the window, removing 24,700 SNPs (5,413 SNPs remaining) within the low error SNP data, and removing 24,422 SNPs (5,688 SNPs remaining) within the high error SNP data. Heavy LD pruning removed SNPs within a 50 SNP window that had $r^2 > 0.09$ with other SNPs (correspond to a VIF > 1.1), removing 28,743 SNPs (1,370 SNPs remaining) within the low error SNP data, and removing 28,732 SNPs (1,378 SNPs remaining) within the high error SNP data. We used VIF LD pruning because we found this procedure led to more consistent SNP densities across different SNP platforms than LD pruning based on pairwise comparisons of SNPs. Using these two levels of LD-pruned SNP data along with the unpruned SNP data, we ran a total of 192 ROH analyses in PLINK (2 autozygosity levels x 2 genotyping error levels x 3 LD-pruning levels x 2 heterozygote allowances x 8 ROH SNP size thresholds), with specific parameters detailed in Table 3.1. We used SNP size ROH thresholds rather than kb length ROH thresholds in PLINK because the former outperformed the latter (data not shown), which is

likely because SNP size thresholds are more robust to the variance in SNP density across the genome.

3.2.5.3 GERMLINE

The principal use of GERMLINE (Gusev et al., 2009) is identity by descent (IBD) mapping between individuals, where ROH analysis is the special case of IBD within an individual. A ROH analysis in GERMLINE is carried out with the -homoz or -homoz-only command. For reasons of efficiency, GERMLINE breaks up SNP data into non-overlapping windows of a user-specified length in SNPs (default is 128 SNPs). Windows that are completely homozygous are tagged. If several tagged windows are in a row and surpass a user-defined length threshold in terms of genetic (cM) or physical (kb) distance, the region is called a ROH. We used minimum genetic distance rather than minimum kb for our ROH thresholds because genetic distance is likely to be more sensitive to variation in recombination rates across the genome. To accommodate various genetic distances, we set the window size threshold to be the expected number of SNPs for a given genetic distance. For example, given that our simulated data encompasses 156 cM, a 1 cM window size would be 193 SNPs in the low error SNP data (30,110 SNPs / 156 cM), but only 9 SNPs in the low error SNP data heavily pruned for LD (1,370 SNPs / 156 cM). Because ROHs must be in multiples of the window size threshold, GERMLINE's resolution of ROH start/end points is less fine grained than PLINK's, and small autozygous segments may be missed by GERMLINE. Like PLINK, GERMLINE also allows for a user-defined number of heterozygous calls to exist in a window (other user-defined parameters are detailed at http://www.cs.columbia.edu/~gusev/germline/). Also like PLINK, GERMLINE does not account for SNP MAF or LD. Thus, we included the same MAF and LD pruned data subsets used in the PLINK analysis. We ran at total of 192 ROH analyses in GERMLINE (2

autozygosity levels x 2 genotyping error levels x 3 LD-pruning levels x 2 heterozygote allowances x 8 ROH cM size thresholds), with specific parameters detailed in Table 3.1.

3.2.5.4 BEAGLE

BEAGLE's ROH detection algorithm (Browning & Browning, 2010) uses a fundamentally different approach than PLINK or GERMLINE. BEAGLE employs a Hidden Markov Model (HMM) that incorporates LD between SNPs and haplotype probabilities from the entire sample when calling ROH segments (for details, see Browning & Browning, 2010). Two user-defined prior probabilities set the baseline expectation of detecting an autozygous segment in a single cM stretch of SNP data. The non-HBD to HBD transition rate is the prior probability per cM of a non-autozygous SNP becoming autozygous (default = 0.0001) (HBD stands for "homozygous by descent," which is conceptually identical to what we term "autozygosity"). Lower values mean that autozygosity is expected to be less common. Conversely, the HBD to non-HBD transition rate is the prior probability per cM of an autozygous SNP becoming non-autozygous (default = 1). Lower values mean that autozygous runs are expected to be shorter. BEAGLE outputs an individual x SNP matrix of posterior probabilities that each SNP is part of an autozygous segment.

Because BEAGLE's HBD program accounts for LD, we did not use pruned SNP data in the BEAGLE analysis. For prior parameters, we set the non-HBD to HBD transition rates of 0.0001, 0.01, and 0.1, and set the HBD to non-HBD transition rates at 1, 0.5, and 0.25. It should be noted that this is a large range of priors, and that they include the default priors (Browning & Browning, 2010). As suggested in Browning and Browning (2010), to avoid false negatives, we also used the maximum posterior probability for each SNP across 10 independent iterations of their program to compare with results from a single iteration. In all, we ran at total of 72

BEAGLE analyses (2 autozygosity levels x 2 genotyping error levels x 9 prior probabilities x 2 iteration levels).

All simulations, statistical programming, and graphing were done using R statistical software 2.11.1 (http://r-project.org).

3.2.6 Comparison of True Autozygosity to Detected ROH

To get an estimate of the type 1 error rates (detecting a ROH that is not autozygous) and type 2 error rates (failing to detect an autozygous ROH) for each program, we compared known autozygosity from the sequence data (as described above) to detected ROH from each analysis. These two types of errors were called on a per-SNP rather than per-ROH basis, as calls can be correct for one part of an autozygous segment and wrong for another. To calculate the type 1 error rate, we summed the total number of SNPs that were type 1 errors and divided it by the total number of non-autozygous SNPs. To calculate the type 2 error rate, we summed the total number of SNPs that were type 2 errors and divided it by the total number of autozygous SNPs. We then estimated *d′*, an index of measurement sensitivity in signal detection that incorporates both type 1 and type 2 error rates (higher *d′* values mean greater sensitivity). *d′* is estimated as: φ(1 - type 2 error rate) - φ(type 1 error rate), where φ is the distribution function of a standard normal, which converts a proportion to a Z-score value. Our estimated *d′* values are included in the table of our top regression power estimates (see *Results*).

3.2.7 Estimation of Regression Power

While *d′* measurements are a good estimate of measurement sensitivity, it has limitations. First, as the type 1 or type 2 error rate approaches zero, *d′* approaches infinity. Second, two identical *d'* estimates can have very different ramifications on the actual number of errors made if the prior probabilities of the errors differ, making it difficult to know which ratio of type 1 to

type 2 error rates is optimal. An alternative and preferable method to *d'* is to ask what ratio of type 1 to type 2 errors would maximize statistical power to detect a relationship between whole-genome autozygosity burden and a phenotype (assuming some base rate of autozygosity). This approach not only circumvents the limitations surrounding *d'* estimates, but also addresses a commonly tested hypothesis in clinical ROH research. Furthermore, power results derived from testing whole-genome ROH burden apply to single ROH association hypotheses (e.g. ROH mapping) as well because the error probabilities in detecting autozygosity are equivalent at the single ROH level and at the whole-genome level.

To estimate statistical power of a whole-genome ROH burden analysis informed by the type 1 and type 2 error rates, we simulated a sample of 2,000 individuals, with every individual's genome split into 'potential' autozygous segments of equal length (7,565 segments for the 20 generation autozygosity map (3,200 Mb/423 kb) and 18,935 segments for the 50 generation autozygosity map (3,200 Mb/169 kb)). Each segment had a probability of being autozygous at the rate observed in our simulated sequence data (0.36% within 20 generations and 0.91% within 50 generations). While the true level of autozygosity in modern outbred populations is unknown, these base rates are likely to be close to the true level because the simulation parameters and selective sampling of the sequence data were chosen to mimic the level of LD and distribution of ROHs found in modern European populations. A continuous phenotype was created for each individual such that the summed autozygous segments within individuals accounted for a small percentage (1%) of the variance of their phenotype score. The choice of the variance accounted for (1% in this case) is purely arbitrary; other choices would raise or lower the absolute levels of detected power but have no effect on the relative differences between various estimates of statistical power, and therefore would have no influence on our final conclusions. We then

simulated true calls, false calls (type 1 errors), true non-calls, and false non-calls (type 2 errors) using the observed error rates from each ROH analysis. We summed the called ROHs for each individual (made up of ROHs that are both true positives and type 1 errors) and regressed the simulated phenotype on this sum. To derive our power estimates, we repeated our simulated regression 1,000 times for each analysis. Regression power was defined as the proportion of trials associated with a positive slope and p-value < 0.05, so a power estimate of 0.5 would mean that 500 of the 1,000 simulations positively associated the simulated phenotype with overall ROH burden.

3.3 Results

Overall, PLINK consistently generated the highest regression power estimates for detecting autozygosity, outperforming GERMLINE and BEAGLE. Figures 3.5 and 3.6 show the range of regression power estimates for PLINK, GERMLINE, and BEAGLE (See figures 3.7, 3.8, 3.9, and 3.10 for type 1 and type 2 error rates). In general, PLINK and GERMLINE power estimates were highly sensitive to their tuning parameters, whereas BEAGLE power estimates were insensitive to all prior probability parameters. For the PLINK results, power was highest when using moderately LD-pruned SNP data, with un-pruned and heavily LD-pruned SNP data performing below the top power estimates. Not surprisingly, higher genotyping error rates generally led to lower regression power estimates in PLINK and GERMLINE.

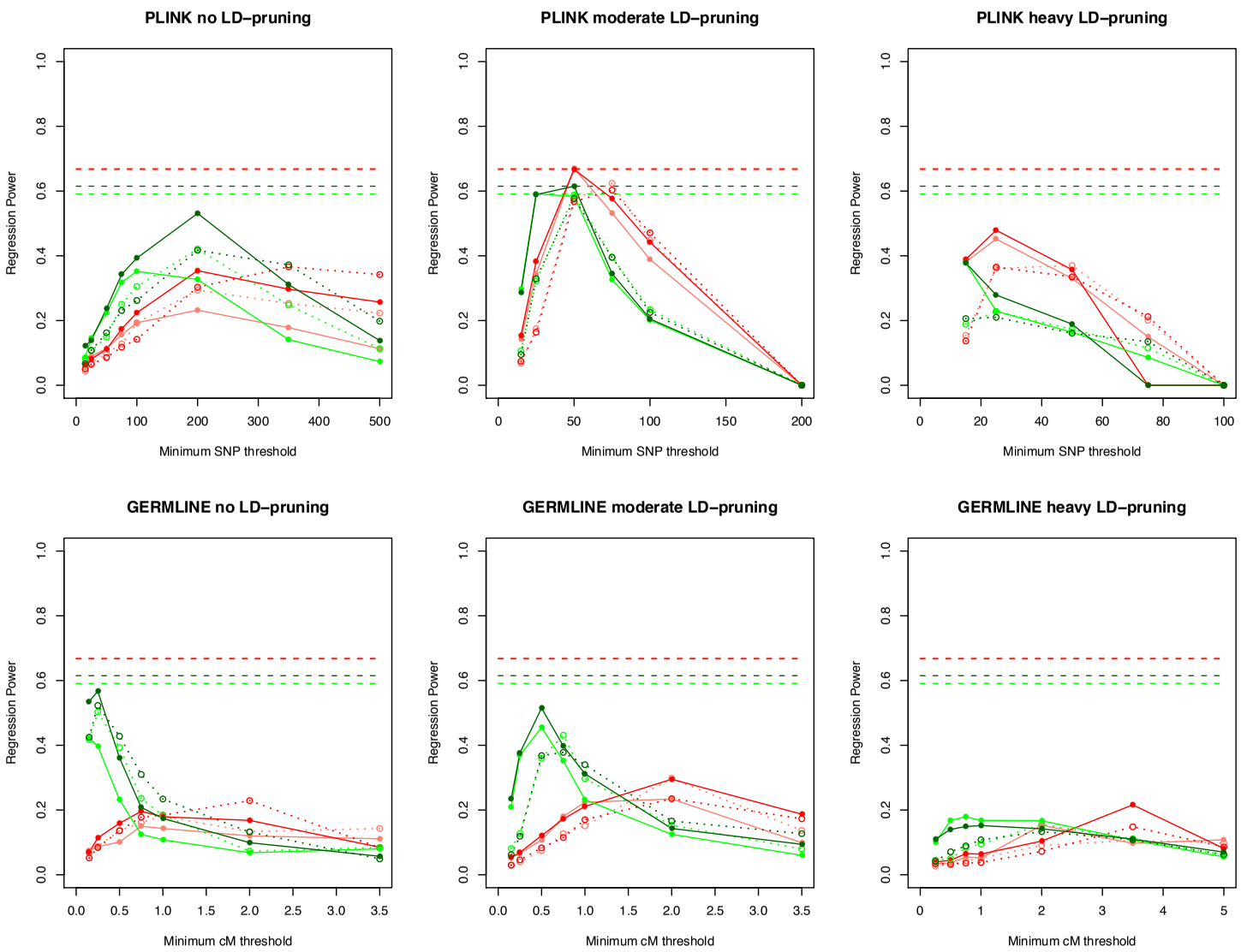

Figure 3.5 Initial Regression Power Results for PLINK and GERMLINE. See Legend for Figure 3.5 for details.

3.3.1 Legend for Figure 3.5

Each graph represents power estimates for each program using unpruned, moderate LD-pruned, or heavy LD-pruned SNP data across different minimum SNP (PLINK) or cM (GERMLINE) lengths. The color of each line represents power estimates with respect to autozygosity within the past 20 and 50 generations, and within low and high genotyping error rates, and are as follows:

Dark red – Autozygosity up to 20 generations and low genotyping error rate

Light red – Autozygosity up to 20 generations and high genotyping error rate

Dark green – Autozygosity up to 50 generations and low genotyping error rate

Light green – Autozygosity up to 50 generations and high genotyping error rate

Power estimates allowing for no heterozygotes are represented by solid lines, whereas estimates allowing for one heterozygote are represented by dotted lines. The horizontal dashed lines represent the initial maximum power estimates with respect to autozygosity within the past 20 and 50 generations, and within low and high genotyping error rates.

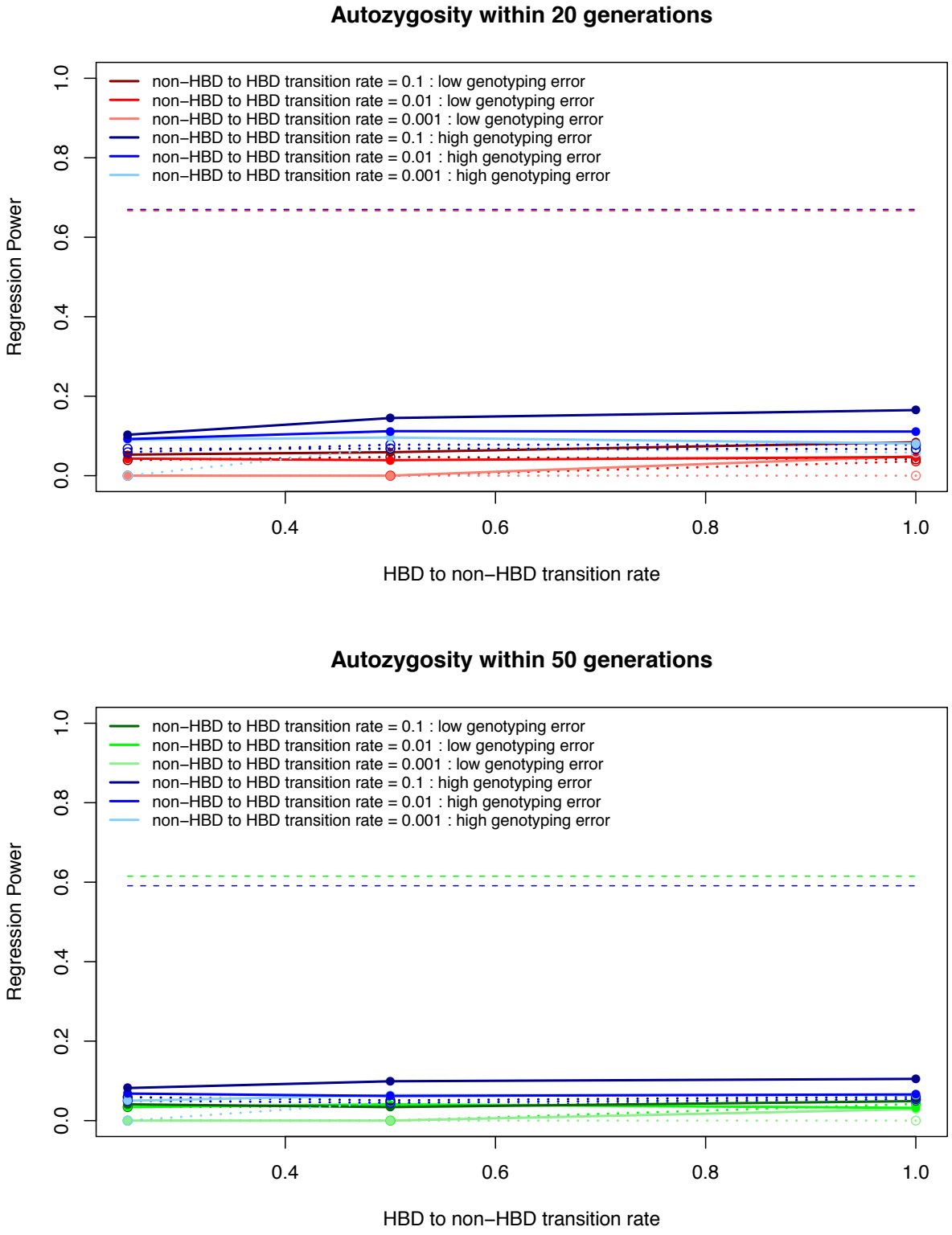

Figure 3.6 Regression Power Results for BEAGLE. See Legend for Figure 3.6 for details.

3.3.2 Legend for Figure 3.6

Each graph represents power estimates with respect to autozygosity within the past 20 and 50 generations. The color of each line represents the non-HBD to HBD transition rate within low and high genotyping error rates, with varying hues reflecting different rates. Solid lines represent power estimates of the maximum posterior probability from 10 BEAGLE iterations, whereas dotted lines represent the estimates from a single BEAGLE iteration. The horizontal dashed lines represent the initial maximum power estimates with respect to autozygosity within the past 20 and 50 generations, and within low and high genotyping error rates.

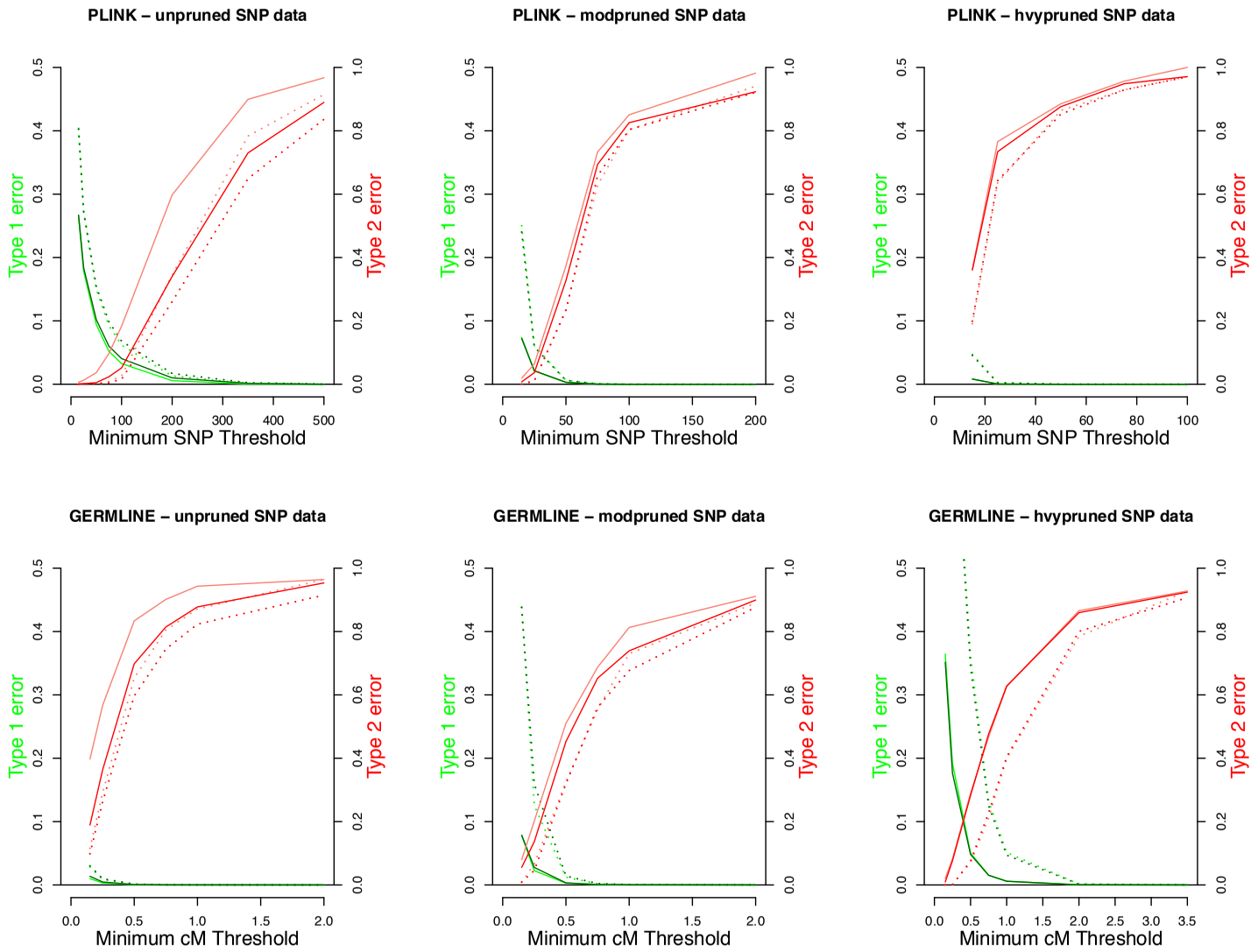

Figure 3.7 Type 1 and Type 2 Errors for PLINK and GERMLINE (Fauto = 20 years). See Legend for Figures 3.7 and 3.8 for details.

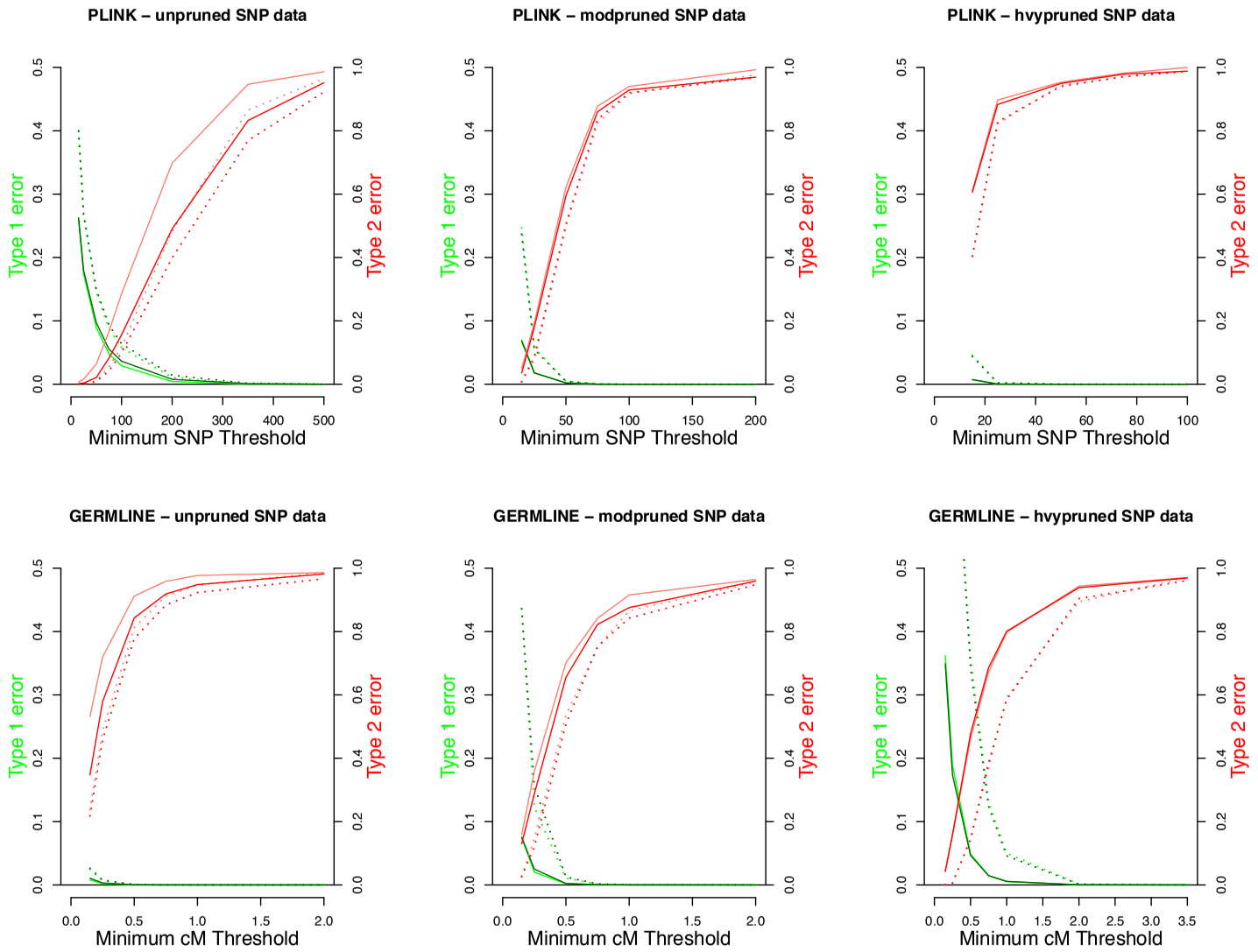

Figure 3.8 Type 1 and Type 2 Errors for PLINK and GERMLINE (Fauto = 50 years). See Legend for Figures 3.7 and 3.8 for details.

3.3.3 Legend for Figures 3.7 and 3.8

Figure 3.7 represents type 1 and type 2 errors to detect autozygosity within the past 20 generations. Figure 3.8 represents type 1 and type 2 errors to detect autozygosity within the past 50 generations. Type 1 and type 2 error rates for each program are shown using unpruned, moderately LD-pruned, or heavily LD-pruned SNP data across different minimum SNP (PLINK) or cM (GERMLINE) lengths. Green lines represent type 1 error rates and are measured along the X-axis, while red lines represent type 2 error rates and are measured along the Z-axis. Color hues are as follows:

Dark green and dark red – low genotyping error rate

Light green and light red – high genotyping error rate

Type 1 and type 2 error rates allowing for no heterozygotes are represented by solid lines, whereas error rates allowing for one heterozygote are represented by dotted lines.

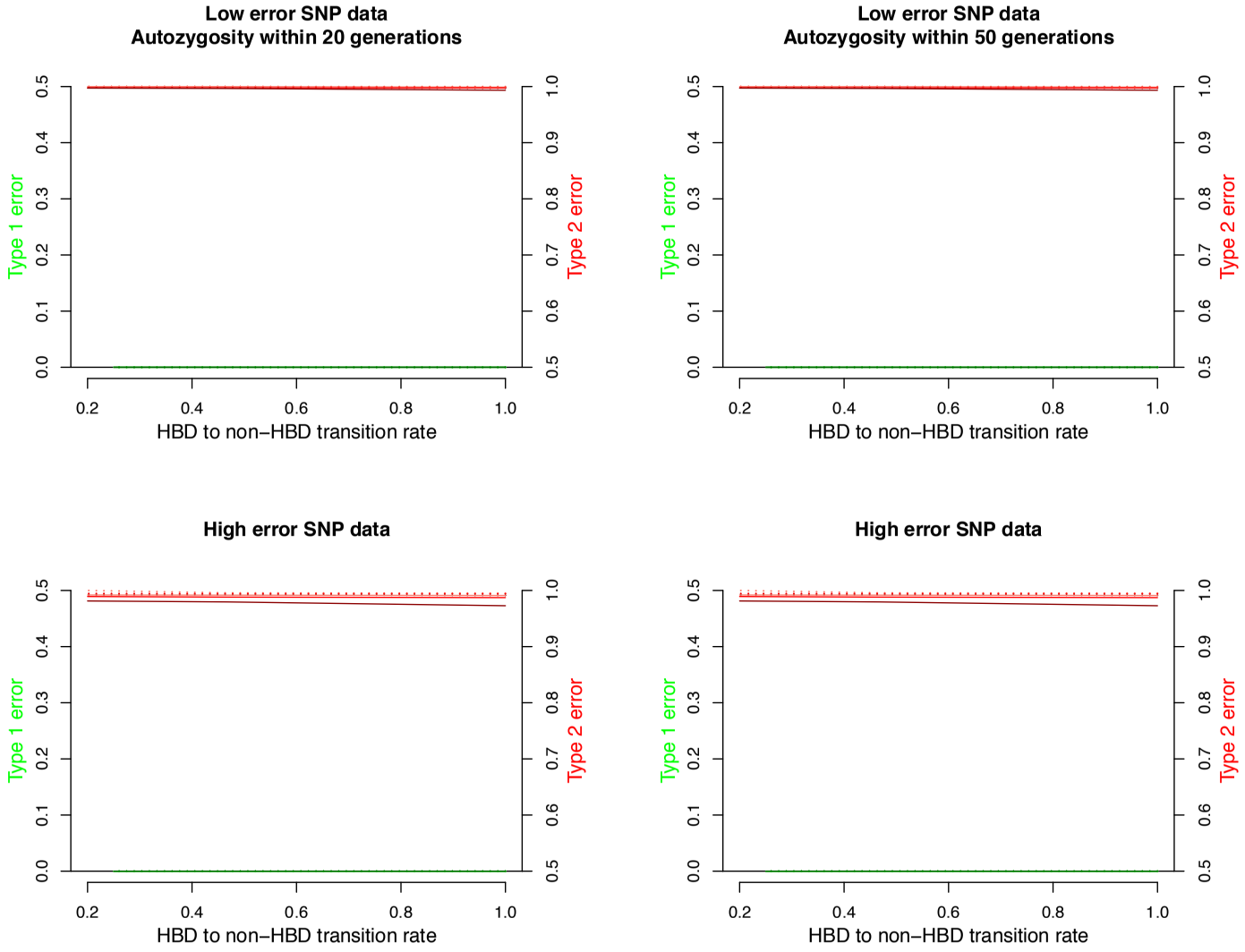

Figure 3.9 Type 1 and Type 2 Errors for BEAGLE. See Legend for Figure 3.9 for details.

3.3.4 Legend for Figure 3.9

Type 1 and type 2 error rates for BEAGLE are shown using for autozygosity within 20 and 50 generations, and within high and low genotyping error rates. Green lines represent type 1 error rates and are measured along the X-axis, while red lines represent type 2 error rates and are measured along the Z-axis. Color hues are as follows:

Dark green and dark red – non-HBD to HBD transition rate of 0.1

Green and red – non-HBD to HBD transition rate of 0.01

Light green and light red – non-HBD to HBD transition rate of 0.001

Solid lines represent type 1 and type 2 error rates from the maximum probability of 10 BEAGLE iterations, whereas dotted lines represent Type 1 and type 2 error rates of a single BEAGLE iteration.

Given these initial results, we decided to look in more detail within the PLINK ROH analyses for the optimum level of LD-pruning and minimum SNP threshold. In addition to moderate LD-pruning, we included "light" LD-pruning, where we removed SNPs that had r2 > 0.9 with other SNPs in a 50 SNP window (VIF > 10), removing 19,662 SNPs (10,451 SNPs remaining) within the low error SNP data, and removing 17,756 SNPs (12,354 SNPs remaining) within the high error SNP data. We also included "moderate-heavy" LD-pruning, where we removed SNPs that had r2 > 0.25 in a 50 SNP window (VIF > 1.33), removing 26,375 SNPs (3,738 SNPs remaining) within the low error SNP data, and removing 26,213 SNPs (3,897 SNPs remaining) within the high error SNP data. We also varied the minimum SNP threshold between 10 and 125 SNPs in increments of 5 SNPs.

The fine-scale results from PLINK (Figure 3.10) show that power was maximized using light to moderate LD-pruning, with the power from moderate-heavy LD-pruning peaking well below the maximum power estimates. Comparisons with the results presented in Figure 3.10 show that both light and moderate LD-pruning are roughly equivalent in terms of regression power. The effect of allowing for a heterozygote call often depended on the minimum length of called ROH, but with respect to the highest power results, allowing for heterozygote calls never performed better than allowing for no heterozygotes. Finally, the optimum SNP threshold for calling ROHs depended on how ancient the autozygosity was. When using moderate LD-pruned data, a 45 SNP minimum threshold worked best to detect autozygosity within the past 20 generations, whereas a 35 SNP minimum threshold worked best to detect autozygosity within the past 50 generations. In general, higher minimum SNP thresholds (and therefore longer detected ROH segments) detect recent autozygosity better, whereas lower thresholds better detect ancient autozygosity. This is expected as autozygous segments are broken into shorter lengths due to

recombination over generational time. Table 3.2 lists the top regression power results within the past 20 and 50 generations for both high and low genotyping error rates.

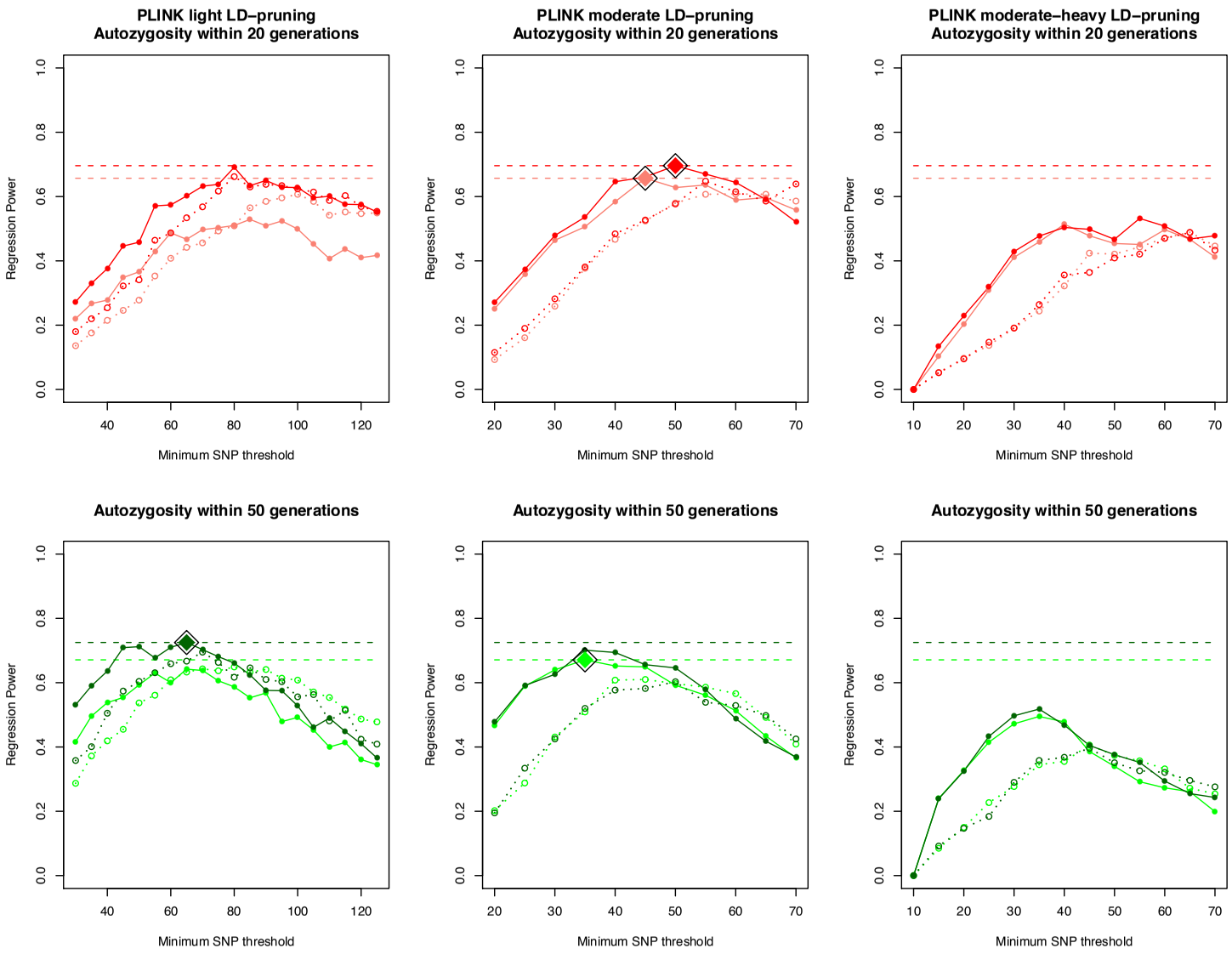

Figure 3.10 Fine Scale Regression Power Results for PLINK. See Legend for Figure 3.10 for details.

3.3.5 Legend for Figure 3.10

Each graph represents power estimates for PLINK using light LD-pruned, moderate LD-pruned, or moderate-heavy LD-pruned SNP data across different minimum SNP (PLINK). The top graphs represent power estimates with respect to autozygosity within the past 20 generations, and the bottom graphs represent autozygosity within the past 50 generations. The color of each line is identical to those used in figure 3, and are as follows:

Dark red – Autozygosity up to 20 generations and low genotyping error rate

Light red – Autozygosity up to 20 generations and high genotyping error rate

Dark green – Autozygosity up to 50 generations and low genotyping error rate

Light green – Autozygosity up to 50 generations and high genotyping error rate

Power estimates allowing for no heterozygotes are represented by solid lines, whereas estimates allowing for one heterozygote are represented by dotted lines. The horizontal dashed lines represent maximum power estimates with respect to autozygosity within the past 20 and 50 generations, and within low and high genotyping error rates, and the large diamond points represent the SNP threshold where PLINK reaches the maximum power estimates.

Table 3.2 Top Regression Power Results

Program	Power	Length	Est kb	Het	α	β	d'	LD-pruning
Autozygosity within 20 generations (low genotyping error)								
PLINK	0.696	50 SNPs	1108	0	0.003	0.33	3.19	Moderate
PLINK	0.691	80 SNPs	919	0	0.003	0.31	3.2	Light
PLINK	0.670	55 SNPs	1219	0	0.002	0.45	3.02	Moderate
PLINK	0.662	80 SNPs	918	1	0.005	0.23	3.28	Light
Autozygosity within 20 generations (high genotyping error)								
PLINK	0.657	45 SNPs	949	0	0.004	0.29	3.19	Moderate
PLINK	0.636	55 SNPs	1160	0	0.002	0.47	2.97	Moderate
PLINK	0.628	50 SNPs	1054	0	0.003	0.37	3.07	Moderate
PLINK	0.608	60 SNPs	1265	1	0.003	0.41	3.01	Moderate
Autozygosity within 50 generations (low genotyping error)								
PLINK	0.725	65 SNPs	746	0	0.005	0.40	2.83	Light
PLINK	0.712	50 SNPs	574	0	0.01	0.23	3.06	Light
PLINK	0.710	60 SNPs	689	0	0.006	0.36	2.88	Light
PLINK	0.709	45 SNPs	517	0	0.012	0.18	3.14	Light
Autozygosity within 50 generations (high genotyping error)								
PLINK	0.671	35 SNPs	738	0	0.007	0.38	2.76	Moderate
PLINK	0.652	40 SNPs	844	0	0.004	0.47	2.69	Moderate
PLINK	0.649	45 SNPs	949	0	0.003	0.55	2.64	Moderate
PLINK	0.649	80 SNPs	777	1	0.006	0.40	2.73	Light

Listed are the top four repression power results for detecting autozygosity within the past 20 and 50 generations, and within low and high SNP genotyping error. Low genotyping error = 0.2% genotyping error rate. High genotyping error = 1% genotyping error rate. **Length** = Minimum SNP or cM length to call a segment. **Est kb** = Minimum expected kb distance for the given length. **Het** = Number of Heterozygotes allowed in a segment. $\boldsymbol{\alpha}$ = Type 1 error rate. $\boldsymbol{\beta}$ = Type 2 error rate. $\boldsymbol{d'}$ = d-prime estimate. **LD-pruning** = Level of LD-pruning. Moderate = Removal of SNPs within a 50 SNP window that had VIF greater than 2 (i.e. --indep 50 5 2 in PLINK). Light = Removal of SNPs within a 50 SNP window that had VIF greater than 10 (i.e. --indep 50 5 10 in PLINK).

Regression power estimates for GERMLINE were similar in pattern but consistently lower than power estimates for PLINK, which was likely driven by the lower start/end resolution of ROH calling in GERMLINE (see *Methods*). To ensure that the discrepancy between GERMLINE and PLINK results were not due to the different ways that ROH length thresholds were defined, we also looked at minimum SNP thresholds in GERMLINE (as opposed to cM length), finding virtually identical results to PLINK, but with slightly lower regression power estimates at all thresholds (data not shown).

BEAGLE was very conservative in detecting autozygous segments, with consistently high type 2 error rates despite using the maximum posterior probability from 10 iterations and a reduced threshold for calling ROH to any posterior probability greater than 10% (a 50% posterior probability threshold was used in Browning & Browning (2010). In short, applying liberal ROH calling parameters did not significantly improve the conservative estimates for BEAGLE.

3.3.6 Computational Time

There were major differences in computational time between the three programs. For our 120 Mb data using 1,000 individuals, a single PLINK and GERMLINE ROH analysis took under 30 seconds, whereas a single BEAGLE HBD analysis took ~150 minutes (about half this time was taken to phase the data - a necessary step in the way their algorithm was written). Both PLINK and GERMLINE analysis times scale up linearly with respect to distance, so a whole-genome analysis for 1,000 individuals should take under 11 minutes. On the other hand, BEAGLE analysis run time scales exponentially, so while BEAGLE runs a separate analysis on each chromosome, chromosome size and sample size greatly affects the computational time.

3.4 Discussion

By simulating sequence and SNP data to match important population genetic properties found in empirical SNP data, we were able to compare the performance of ROH detection programs to identify true levels of autozygosity expected in large, outbred populations of European heritage. PLINK consistently outperformed both GERMLINE and BEAGLE, producing higher regression power estimates for detecting autozygosity within 20 and 50 generations, regardless of the genotyping error rate. While we found, as expected, that GERMLINE performed worse than PLINK due to the lower resolution of start/end points of ROHs, we were surprised by the lower performance of BEAGLE. Specifically, we expected that the incorporation of MAF and LD information in BEAGLE's hidden markov model to result in higher accuracy to detect autozygosity. However, this did not appear to be the case. Rather, BEAGLE was consistently conservative, and changing the tuning parameters did not alter this. As currently designed, at least, BEAGLE appears optimized to detect rather long autozygous segments arising from a recent common ancestor, but has not been designed to detect the more ambiguous autozygous signals from distant common ancestors.

3.4.1 Recommendations

Our results suggest that PLINK is the most suitable program for detecting autozygosity arising from distant ancestors (see Table 3.3). Our results also demonstrate that removing low MAF (< 0.05) SNPs and removing SNPs through light-to-moderate LD-pruning (e.g., VIF between 2 and 10) prior to the analysis minimizes the trade-off between the exclusion of non-autozygous ROHs and the cost of missing shorter autozygous ROHs. In particular, LD-pruning improves detection accuracy by removing redundant SNPs within SNP-dense regions, making SNP coverage more uniform with respect to recombination distance and allowing ROH calls to

be less dependent on the variation in SNP density across platforms. Moreover, our results suggest not allowing any heterozygote SNPs to exist in a called ROH. Our recommendations regarding the minimum SNP threshold, however, vary slightly depending on what strength of LD pruning is used and on the time since the common ancestor of the autozygous segment. For example, to detect autozygosity arising from a common ancestor within the past 50 generations, a 65 SNP minimum threshold is preferred when using light LD-pruning, while a 35 SNP minimum threshold is preferred when using moderate LD-pruning. Both thresholds cover a minimum physical distance of ~750 kb, but moderate LD-pruning retains fewer SNPs across that distance. Analyses geared towards detecting more recent autozygosity should increase the minimum SNP length threshold. When we examined autozygosity arising from common ancestors within the past 20 generations, a 45 to 50 SNP minimum threshold performed best when using moderate LD-pruning, depending on how much error is expected in the SNP data. Because our recommendations include LD-pruning, the increased SNP density of modern platforms should have a minimal effect on our recommendations, as the level of LD between SNPs remains roughly constant regardless of SNP density.

Table 3.3 Top Recommendations for Detecting Autozygosity

Autozygosity Detection	**SNP data**	**Top Program**	**Chosen parameters**
Within the past 20 generations	Low genotyping error	PLINK	- Moderate LD-pruning - 50 SNP threshold - No heterozygote allowed
Within the past 20 generations	High genotyping error	PLINK	- Moderate LD-pruning - 45 SNP threshold - No heterozygote allowed
Within the past 50 generations	Low genotyping error	PLINK	- Light LD-pruning - 65 SNP threshold - No heterozygote allowed
Within the past 50 generations	High genotyping error	PLINK	- Moderate LD-pruning - 35 SNP threshold - No heterozygote allowed

Low genotyping error = 0.2% genotyping error rate. High genotyping error = 1% genotyping error rate. Moderate LD-pruning = Removal of SNPs within a 50 SNP window that had VIF greater than 2 (i.e. --indep 50 5 2 in PLINK). Light LD-pruning = Removal of SNPs within a 50 SNP window that had VIF greater than 10 (i.e. --indep 50 5 10 in PLINK).

3.4.2 Limitations

Despite our efforts to simulate realistic sequence and SNP data, additional factors that occur in real data, such as genotyping plate effects, autozygous runs caused by positive selection, hemizygous deletions, SNP-poor centromeres, and uniparental isodisomy (two copies of the same chromosome from a single parent) were not simulated and may affect estimates of ROH in real data. We did not investigate several additional questions, such as the effects of our recommendations for detecting ROHs in non-European heritage or isolated populations. We also did not investigate optimal thresholds/program for detecting more recent or more ancient autozygosity. However, autozygosity arising from more recent ancestors becomes increasingly easy to detect, and most programs/thresholds should detect it with very high fidelity. Autozygosity arising from more ancient common ancestors is more difficult to detect but may nevertheless be detectable as SNP chips become denser. However, the variation between individuals in overall burden of such ancient autozygosity becomes very small (Keller et al., 2011), and thus there are diminishing returns from attempting to detect such ancient autozygosity, at least with respect to analyses investigating overall ROH burden.

3.4.3 Conclusion

PLINK has been the most commonly used ROH detection program to date. However, only one study analyzed data that was pruned for LD (Spain et al., 2009), which we have found to be an important step for improving the accuracy of detecting autozygous ROHs. Two studies adjusted their minimum ROH SNP threshold upward to account for LD creating ROH by chance (Enciso-Mora et al., 2010; Hosking et al., 2010), but our results show that without explicit LD-pruning, the increase in detection error cannot be overcome by larger ROH size thresholds. While thresholds for calling ROHs varied across previous studies due to the lack of consensus

criteria, most previous studies adopted more liberal thresholds than our results suggested are optimal. Thus, power in previous studies was likely to be lower than optimal due to inclusion of many non-autozygous ROHs.

The current study is the first of its kind to directly assess the ability of current ROH detection programs to estimate genome-wide autozygosity. Our results should apply equally well to research on autozygosity using whole-genome ROH burden or single ROH association mapping. None of the programs tested perfectly detects autozygosity, and new parameters and algorithms may further minimize detection error and increase sensitivity to detect autozygous segments. Until then, the results in this study represent an important step for developing working 'consensus criteria' for defining ROH.

ACKNOWLEDGEMENTS

DP Howrigan's contribution to this work was partially supported by two institutional training grants from the National Institute of Child Health and Human Development (T32 HD007289, Michael C. Stallings, Director) and the National Institute of Mental Health (T32 MH016880, John K. Hewitt, Director) awarded to the Institute for Behavioral Genetics, University of Colorado. JW Smoller's contribution to this work was supported in part by the National Institute of Mental Health (MH063445). Funding support for the Molecular Genetics of Schizophrenia (MGS) genome-wide association of schizophrenia study was provided by the National Institute of Mental Health (NIMH) (R01 MH67257, R01 MH59588, R01 MH59571, R01 MH59565, R01 MH59587, R01 MH60870, R01 MH59566, R01 MH59586, R01 MH61675, R01 MH60879, R01 MH81800, U01 MH46276, U01 MH46289 U01 MH46318, U01 MH79469, and U01 MH79470), and the genotyping of samples was provided through the Genetic Association Information Network (GAIN) and the MGS non-GAIN network. The datasets used for the analyses described in this manuscript were obtained from the database of genotype and phenotype (dbGaP) found at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession numbers phs000021.v3.p2 and phs000167.v1.p1. Samples and associated phenotype data for the genome-wide association of schizophrenia study were provided by the Molecular Genetics of Schizophrenia Collaboration (PI: P.V. Gejman, North Shore University Health System (NUH)) and University of Chicago (Evanston, Illinois, USA).

Chapter 4

The Effect of Genome-Wide Autozygosity on General Cognitive Ability

4.1 Background

General intelligence, traditionally measured through IQ, is a well-studied quantitative trait that correlates strongly with almost all measures of cognitive ability and maintains strong consistency across different IQ measurements (Carroll, 1993; Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004; Johnson, Nijenhuis, & Bouchard, 2008). From a genetic standpoint, decades of behavioral genetic research on IQ have shown moderate to high heritability estimates (see Erlenmeyer-Kimling & Jarvik, 1963 and Bouchard & McGue, 1981 for reviews) and consistency of heritability across development (Deary et al., 2012; Deary, Johnson, & Houlihan, 2009), both of which suggest a robust genetic component underlying individual differences in IQ. From a practical standpoint, there are salient reasons to elucidate the genetic underpinnings of IQ, as it is found to be a strong predictor of many life outcomes, such as health, longevity, social mobility, and occupational success (Batty, Deary, & Gottfredson, 2007; Deary, 2012).

Recent research has found that the genetic variation in general intelligence is highly polygenic; that is, no single gene or small set of genes has a significant effect (Butcher, Davis, Craig, & Plomin, 2008; Cirulli et al., 2010; Davis et al., 2010; Need et al., 2009), whereas the accumulation of small genetic effects across the genome explain a substantial proportion (40-50%) of the phenotypic variance of IQ (Chabris et al., 2012 ; Davies et al., 2011). Genome-wide effects of common SNPs under a polygenic inheritance framework, however, are unable to infer the contribution of non-additive recessive genetic variation, as the effects of recessive or partially

recessive variants are not fully accounted for. One potential driver of non-additive variation is through directional dominance, where selection has purged out additive deleterious variants, leaving a higher pool of recessive variants in the population than expected (Morton, 1978). As a result, the effects of directional dominance can be revealed through inbreeding depression, where the offspring of closely genetic relatives show higher rates or disorder and lower scores on fitness-related traits due to the expression of recessive variants. The detrimental effects of inbreeding depression have been well documented in non-human literature (DeRose & Roff, 1999), form the basis of many rare familial disorders, and have been found in some complex disorders (Lebel & Gallagher, 1989; Rudan et al., 2003; Rudan et al., 2004). With respect to IQ, studies have shown that offspring of consanguineous marriages have lower IQ scores than the general population, supporting the role of inbreeding depression in general intelligence (Afzal, 1988; Afzal, 1993; Agrawal, Sinha, & Jensen, 1984; Morton, 1978; Woodley, 2009).

4.1.1 Autozygosity and *Froh*

At the genetic level, inbreeding depression is a result of autozygosity, where stretches of two homologous chromosomes in the same individual are identical by descent. While highly inbred individuals will have a substantial proportion of their genome as autozygous (e.g. first cousin inbreeding will lead to 6.25% genome-wide autozygosity on average), autozygosity can still occur in outbred populations, albeit at lower levels as the common ancestor of an autozygous stretch can go back dozens of generations. With the advent of high-density SNP arrays, recent research has shown that detecting Runs of Homozygosity (ROH) on genome-wide SNP arrays provides the most informative measure of autozygosity, outperforming the inbreeding coefficient *F* (a measure of how overall homozygosity rates compare to Hardy-Weinberg expectations) and known pedigrees (Carothers et al., 2006; Keller et al., 2011). Genome-wide autozygosity burden,

termed here as *Froh*, is calculated by determining the percent of the genome made up of autozygous ROH.

To date, a number of studies have examined the effect of *Froh* burden and individual ROH regions on case/control and quantitative phenotypes, finding only a few significant results (see Ku, Naidoo, Teo, & Pawitan, 2011 for a review). However, recent simulation work has shown that the small amount of variation of autozygous ROH in outbred populations requires sample sizes ranging from 20,000 to 60,000 to achieve adequate power when testing the relationship of distant inbreeding to phenotypic variation (Keller et al., 2011).

4.1.2 Overview of Current Study

With this in mind, the current study has drawn together genotype and IQ data from nine separate datasets, combining to a total sample size of 4,360 individuals. Despite being well below the suggested sample size, this is the first study to date that assesses the relationship of IQ to autozygosity using dense genome-wide SNP data, and at the very least can indicate whether or not there is a trend towards increasing autozygosity associating with lower IQ. Briefly, the current study assesses the prediction that increased autozygosity burden, or *Froh*, will associate with lower IQ scores across the normal range. In addition, effects of autozygosity frequency and length are also assessed to see if more recent and/or rare autozygous segments associate with lower IQ scores. Genome-wide ROH mapping is then examined to see if specific regions of the genome associate with IQ when autozygous. Finally, a new method is introduced whereby ROH mapping regions are rank-ordered by significance and tested for an excess of positive or negative effects, with promising subsets of regions examined through cross-validation to see if *Froh* burden restricted to a smaller subset of the genome can improve IQ prediction.

4.2 Methods

Data were ascertained from five separate projects conducted across five geographical regions: the GAIN International Multi-Center ADHD Genetics Project (Northern Europe, British Islands, Spain), the Brisbane Adolescent Twin Study (Australia), the Lothian Mental Survey (Scotland), the Manchester and Newcastle longitudinal studies of cognitive aging (England), and the IBG study (Colorado, USA). All subjects in the final samples were of European descent. Descriptive statistics for IQ, age, and sex for each sample are presented in Table 4.1.

Table 4.1 Descriptive Statistics of IQ, Age, and Sex Across Datasets.

Dataset	N	Region	IQ measures	Mean Age (SD)	Male (%)	Female (%)
IBG	301	Colorado, USA	WAIS-III: 2 subtests (Ages 16+) WISC-III: 2 subtests (Ages 8-16)	15.91 (1.53)	232 (77%)	69 (23%)
GAIN NE	357	Northern Europe	WAIS-III: 4 subtests (Ages 16+) WISC-III: 4 subtests (Ages 5-16)	10.95 (2.57)	305 (85%)	52 (15%)
GAIN UK	183	United Kingdom	WAIS-III: 4 subtests (Ages 16+) WISC-III: 4 subtests (Ages 5-16)	11.67 (2.83)	165 (90%)	18 (10%)
GAIN SP	68	Spain	WAIS-III: 4 subtests (Ages 16+) WISC-III: 4 subtests (Ages 5-16)	9.40 (2.53)	62 (91%)	6 (9%)
LBC 1921	495	Scotland	Moray House Test Ravens Progressive Matrices Verbal Fluency & Logical Memory National Adult reading Test	79.05 (0.57)	202 (41%)	293 (59%)
LBC 1936	939	Scotland	Moray House Test WAIS-III: 6 subtests National Adult reading Test	69.55 (0.85)	482 (51%)	457 (49%)
MANC	763	England	Cattell Culture Fair Test	64.9 (6.14)	226 (30%)	537 (70%)
NEWC	717	England	Cattell Culture Fair Test	65.71 (6.10)	206 (29%)	511 (71%)
QIMR	537	Australia	MAB: 5 subtests	18.44 (2.49)	255 (47%)	282 (53%)
Total	4360	---	---	51.02 (26.20)	2135 (49%)	2225 (51%)

4.2.1 GAIN IMAGE Project

4.2.1.1 Sample

Data from the IMAGE samples consisted of 958 parent-child trios (n=2,803) from the from the International Multisite ADHD Genetics project, which has the goal of associating SNP markers with ADHD. European Caucasian subjects were recruited from 12 specialist centers in Northern Europe (Belgium [n=108], Germany [n=325], Holland [n=926], and Switzerland [n=77]), the British Isles (United Kingdom [n=393] and Ireland [n=257]), Spain [n=209], and Israel [n=508]. No IQ data was collected in Israel and so it was not analyzed in this study. Entry criteria for probands included: (a) clinical diagnosis of DSM-IV combined subtype ADHD; (b) IQ over 70; (c) aged 6-17 years; (d) one or more sibling(s) in the same age range; (e) both parents available to provide DNA sample or one parent available plus two or more siblings; (f) free of single-gene disorders known to be associated with ADHD; (g) free of neurological disease and damage; (h) living at home with at least one biological parent and one full sibling; and (i) not meeting criteria for autism or Asperger's syndrome. The age range for probands was 5 to 19 years (mean=11, s.d.=2.7). Ethical approval for the study was obtained from the National Institute of Health registered ethical review boards for each center, and informed consent was obtained from all parents and most children. From this initial sample, IQ and genotype data were available from 626 probands. The sample was split into three separate datasets reflecting data collected in Northern Europe, the British Isles, and Spain.

4.2.1.2 IQ Scales

Four subtests (Vocabulary, Similarities, Block Design, and Picture Completion) from either the WISC-III (Wechsler & Corporation, 1991) or the WAIS-III (Wechsler & Corporation, 1997) (depending on subject’s ages) were used to estimate full-scale IQ. These subtests correlate

between .90 and .95 with Full-scale IQ. All IQ scores were corrected for participants' age (mean=100.7, s.d.=15.7).

4.2.1.3 Genotyping

Genomic DNA was isolated from whole blood and genotyped on the Perlegen 600K array. All data was downloaded, with permission, from the publicly available database for Gentoypes and Phenotypes (dbGaP). 438,492 SNPs were available from the initial SNP filtering (QC) implemented on the data. Of note, none of the filters implemented in the initial SNP set were stricter than the QC parameters implemented in the current analysis.

4.2.2 Brisbane Adolescent Twin Study

4.2.2.1 Sample

The Brisbane Adolescent Twin Study sample consists of 3,899 Australian monozygotic and dizygotic twins and twin siblings from 1,324 nuclear families (Wright & Martin, 2004). Twins were recruited in 1992 and again in 1998 through primary and secondary schools in South East Queensland, Australia as well as through the Australian Twin Registry (Hopper, 2002). From this initial sample, IQ and genotype data were available for 1,005 individuals. An individual was chosen from each family by retaining the sibling with fewest missing genotypes was chosen, resulting in 540 unrelated individuals for analysis. Participants were 12-25 years old (mean=18.44, s.d.=2.49) when they took the IQ battery. Subjects who had a history of significant head injury, neurological or psychiatric illness, substance dependence, or were currently taking long-term medications with central nervous system effects were excluded. Informed consent was collected from participants and their parents, and all procedures and protocols were reviewed and approved by the QIMR Human Research Ethics Committee (for additional details, see Wright & Martin, 2004).

4.2.2.2 IQ Scales

Participants came to the lab for in-person assessments four times; full-scale IQ was assessed using a shortened version of the Multidimensional Aptitude Battery, which included three verbal subtests (Information, Arithmetic, Vocabulary) and two Performance subtests (Spatial and Object Assembly). The MAB was patterned after the WAIS-R and, as a result, possesses good psychometric properties (Gignac, 2006; Jackson, 1998).

4.2.2.3 Genotyping

Genomic DNA was derived from whole blood and genotyped on the Illumina 610-Quadv1 array. Genotype calls were made using the BRLMM algorithm, resulting in 559,712 SNPs available for analysis.

4.2.3 Lothian Mental Survey

4.2.3.1 Sample

Participants in the Lothian Mental Survey were part of the 1921 Lothian Birth Cohort (n=550) and the 1936 Lothian Birth Cohort (n=1091) from the Scottish Mental Surveys and follow-up studies (Deary, Whiteman, Starr, Whalley, & Fox, 2004; Harris et al., 2007). The following inclusion criteria were applied: (a) cognitive ability scores were available in the late age cohort; (b) there was no history of dementia; (c) Mini-Mental State Examination (Crum, Anthony, Bassett, & Folstein, 1993) score was 24 or greater; (d) SNP genotyping was successful. From this initial sample, IQ and genotype data were available for 526 individuals in the 1921 Lothian Birth Cohort and 1050 individuals in the 1936 Lothian Birth Cohort. Age ranges are detailed below in the description of IQ scales. For both Lothian Birth Cohorts, ethical approval was obtained from the Lothian Research Ethics Committee. Additional approval for the 1936 Lothian Birth Cohort was obtained from Scotland's Multicenter Research Ethics Committee.

4.2.3.2 IQ Scales

Details of cognitive scales have been described elsewhere (Deary et al., 2004; Harris et al., 2007). In both samples, IQ was examined around age 11 as well as around age 79 (1921 Lothian birth Cohort) and around age 70 (1936 Lothian Birth Cohort). However, because age is controlled for as a covariate in all our analyses, only scores from the more extensive IQ examination at older ages were used. Participants in the 1921 Lothian Birth Cohort completed a battery of cognitive tests: The Moray House Test (no 12), Ravens Standard Progressive Matrices (ref 19 in Davies), Verbal Fluency (ref 20) and logical Memory (ref 21), and the National Adult Reading Test. Participants in the 1936 Lothian Birth Cohort completed a similar battery of cognitive tests: The Moray House Test (no 12), a subset of WAIS-III tests consisting of Digit Symbol Coding, Block Design, Matrix Reasoning, Digit Span Backwards, Symbol Search, and Letter Numbering, as well as the National Adult Reading Test.

4.2.3.3 Genotyping

Genomic DNA was isolated from whole blood and genotyped on the Illumina 610-Quadv1 array. Genotype calls were made using the BRLMM algorithm, resulting in 599,011 SNPs available for analysis.

4.2.4 Longitudinal Cognitive Aging Cohorts

4.2.4.1 Sample

From 1983 to 2003, 6,063 participants from Greater Manchester and Newcastle and were recruited by the Manchester Age and Cognitive Performance Research Center to study longitudinal changes in the cognitive function of older adults. Over this period, participants took the same battery of cognitive tests on two separate occasions, which have been averaged for the current study. IQ and genotype data were available for 860 unrelated individuals in the Greater

Manchester sample, and 840 unrelated individuals in the Newcastle sample. Participant age was also averaged for both testing occasions, with an age range of 41-86 years old (mean=65.25, s.d.=6.31). Ethical approval was obtained from the University of Manchester.

4.2.4.2 IQ Scales

Details of cognitive scales have been described elsewhere (Rabbitt et al., 2004). For the current study, scores from the Cattell and Cattell Culture Fair Intelligence tests were used to measure IQ, which correspond to the fluid measure on intelligence. While measures of crystallized intelligence were not used from the IQ battery, it is worthwhile to note that all IQ measures have been standardized within dataset, so the heterogeneity of measurements should be unbiased.

4.2.4.3 Genotyping

Genomic DNA was isolated from whole blood and genotyped on the Illumina 610-Quadv1 array. Genotype calls were made using the BRLMM algorithm, resulting in 599,011 SNPs available for analysis. The values are the same as the Lothian Birth Cohort samples, as genotyping and SNP calling metrics were conducted at the same site (Davies et al., 2011).

4.2.5 CADD Project

4.2.5.1 Sample

Participants in this sample are a subset from the Colorado Center on Antisocial Drug Dependence (CADD). None of the participants comprised the clinically ascertained cohort of the CADD sample, but include participants in subsets of the CADD from the Colorado Twin Study (CTS), Longitudinal Twin Study (LTS), and Colorado Adoption Project (CAP). Further details on the ascertainment and informed consent of the CADD sample is found in Rhea, Gross, Haberstick, & Corley (2006). After ensuring all individuals included were unrelated, IQ and

genotype data were available for 301 individuals. Participants were 12-19 years old (mean=15.91, s.d.=1.53) when they took the IQ test.

4.2.5.2 IQ Scales

For individuals age 16 or older, IQ was assessed using the WAIS-III (Wechsler & Corporation, 1997), whereas individuals under of the age of 16 were assessed using the WISC-III (Wechsler & Corporation, 1991). IQ scores were derived from a combination of the vocabulary and lock design subtests.

4.2.5.3 Genotyping

Genomic DNA was isolated from whole blood DNA and buccal call DNA, and genotyped on the Affymetrix 6.0 array. Due to low genotyping quality and a mixture of blood and buccal cell DNA, genotype calls were made using the Birdseed algorithm, with call intensities being normalized within blood and buccal DNA subsamples. These calls were further refined using Beaglecall (Browning & Browning, 2010), which uses LD information from the entire sample to assisting genotyping QC, resulting in 791,388 SNPs available for analysis. One unique characteristic of Beaglecall is the iterative process of validating genotype calls, which at the final state leads to no missing genotype data as SNPs are either removed or imputed throughout the process. With this information, a genotype by individual SNP missingness interaction was added to the regression model to control for the difference properties of Beaglecall to BRLMM calling algorithms.

4.2.6 Genetic and Sample Quality Control Procedures

Quality control (QC) procedures focused on properties that would be appropriate across different genotyping platforms that differed in SNP density. The main goal—analyzing runs of homozygosity to infer autozygosity—differed from the usual goal of finding associations

between individuals SNPs and a phenotype, and so the procedures adopted were somewhat modified. Because so many SNPs (70-75% depending on the sample) were removed due to linkage disequilibrium pruning during ROH detection (see below), stringent preliminary SNP-cleaning procedures were adopted, as a SNP lost due to cleaning was likely to be in linkage disequilibrium with a nearby SNP. On the other hand, there are also concerns regarding the removal of SNPs or subjects that might be informative with respect to the questions at hand. For example, individuals with excess genome-wide homozygosity are quite informative in the sample, as autozygous segments significantly contribute to observed homozygosity above and beyond Hardy-Weinberg expectations. As a result, participants with excess homozygosity were not removed and Hardy-Weinberg *p*-value thresholds on individual SNPs were relaxed.

The quality control procedures and numbers of individuals or SNPs lost at each step can be found in Table 4.1. Most steps are self-explanatory, so only those needing clarification are discussed. Individuals whose self-reported gender was discrepant from their genotypic sex (as judged using PLINK's *check-sex* command) were dropped, as these individuals might represent sample mix-ups rather than single incorrect responses. Because both homozygosity and phenotypic measures might differ between ethnicities or across different levels of genetic admixture, two procedures were used to ensure ethnic homogeneity within each of the geographically distinct samples. First, individuals were dropped who self-identified as non-Caucasian. Second, individuals who stood out on a measure representing the posterior likelihood of being Caucasian were also dropped. Within each sample, ~ 50,000 SNPs in linkage-equilibrium were selected and merged with HapMap2 reference panels (Frazer et al., 2007). To assess population stratification, PLINK's multi-dimensional scaling command was used to distinguishing Caucasian, Asian, and Yoruban ancestry (based on HapMap allele frequencies).

The first two principle components were drawn from the muti-dimensional scaling assessment, and individuals greater than 10% away from the center of the Caucasian cluster towards the Asian or Yoruban sample were removed from further analyses.

Table 4.1 Genotype Quality Control Procedures

Dataset	GAIN Ne	%	Remaining	GAIN Br	%	Remaining	GAIN Sp	%	Remaining	LBC 21	%	Remaining	LBC 36	%	Remaining	Manc	%	Remaining	Newc	%	Remaining	QIMR	%	Remaining	IBG	%	Remaining
TOTAL SNPs	**438492**			**438492**			**438492**			**506283**			**506283**			**599011**			**599011**			**559712**			**791388**		
TOTAL SAMPLES	**365**			**186**			**69**			**500**			**966**			**860**			**840**			**14160**			**1760**		
Samples																											
A1) Drop discrepant sex	0	0	365	0	0	186	0	0	69		0	500		0	966	1	0.001	859	1	0.001	839	0	0	14160	24	0.014	1736
A2) Drop MZ cotwin missing most genotyping data	0	0	365	0	0	186	0	0	69		0	500		0	966	0	0	859	0	0	839	462	0.033	13698	0	0	1736
A3) Drop individs/families with no info on IQ, sex, or age	0	0	365	0	0	186	0	0	69		0	500		0	966	34	0.04	825	34	0.04	805	13158	0.929	540	1309	0.744	427
A4) Drop self-identified non-Caucasian	0	0	365	0	0	186	0	0	69		0	500		0	966	4	0.005	821	0	0	805	0	0	540	104	0.059	323
A5) Remove ethnic outliers & related genotypes (computed earlier)	0	0	365	0	0	186	0	0	69		0	500		0	966	20	0.023	801	30	0.036	775	3	2E-04	537	22	0.013	301
A6) Drop individuals missing > 2% SNP calls	8	0.022	357	3	0.016	183	1	0.014	68	5	0.01	495	27	0.028	939	38	0.047	763	58	0.075	717	0	0	537	0	0	301
SNPs																											
B1) Drop SNPs on X, Y, and mitochondria	8451	0.019	430041	8451	0.019	430041	8451	0.019	430041		0	506283		0	506283	16420	0.027	582591	16420	0.027	582591	13850	0.025	545862	36857	0.047	754531
B2) Drop HWE < .001	320	7E-04	429721	327	7E-04	429714	186	4E-04	429855	411	8E-04	505872	483	1E-03	505800	1769	0.003	580822	1680	0.003	580911	461	8E-04	545401	452	6E-04	754079
B3) Drop call rate < .995	141167	0.322	288554	179567	0.41	250147	104907	0.239	324948	24482	0.048	481390	30944	0.061	474856	55324	0.092	525498	58267	0.097	522644	39368	0.07	506033	0	0	754079
B4) Drop MAF < .05	11094	0.025	277460	12790	0.029	237357	15272	0.035	309676	2832	0.006	478558	1372	0.003	473484	61026	0.102	464472	61437	0.103	461207	27521	0.049	478512	177063	0.224	577016
B5) Overlapping QC failed SNPs	3535	0.013	280995	5510	0.023	242867	3054	0.01	312730	180	4E-04	478738	175	4E-04	473659	5590	0.012	470062	5406	0.012	466613	302	6E-04	478814	74	1E-04	577090
Genotype information																											
D1) Get pre QC missingness	Done		NA	Done		NA	Done		NA	Done		NA	Done		NA	Done		NA	Done		NA	Done		NA	Done		NA
D2) Get post QC inbreeding coefficient	Done		NA	Done		NA	Done		NA	Done		NA	Done		NA	Done		NA	Done		NA	Done		NA	Done		NA
D3) Get 10 PCA's	Done		NA	Done		NA	Done		NA	Done		NA	Done		NA	Done		NA	Done		NA	Done		NA	Done		NA
LD Pruning																											
E1) Moderate LD-pruning (VIF > 2 in 50 SNP window)	181651	0.646	99344	154598	0.637	88269	229007	0.732	83723	362096	0.756	116642	356350	0.752	117309	351996	0.749	118066	349060	0.748	117553	361911	0.756	116903	477591	0.828	99499
E2) Light LD-pruning (VIF > 10 in 50 SNP window)	82777	0.295	198218	68854	0.284	174013	138132	0.442	174598	230943	0.482	247795	225104	0.475	248555	219711	0.467	250351	218623	0.469	247990	230161	0.481	248653	370318	0.642	206772

4.2.7 Runs of Homozygosity (ROHs) Calling Procedures

ROH were called in PLINK using the --homozyg command (Purcell et al., 2007), which has been found to outperform other programs in accurately identifying autozygous segments (Howrigan, Simonson, & Keller, 2011). The current analysis incorporated the four ROH tuning parameters recommended in (Howrigan et al., 2011), pertaining to distant and recent autozygosity as well as predicted high and low genotyping error rate. In general, each dataset was pruned for either moderate LD (removing ay SNP with R-square > 0.5 with other SNP in a 50 SNP window) or strong LD (removing ay SNP with R-square > 0.9 with other SNP in a 50 SNP window), and the minimum SNP length threshold varied from 35 to 65 SNPs. Additional details on ROH calling recommendations are described in Howrigan et al. (2011). In addition to accurately calling autozygous ROH, determining the frequency of ROHs also has important implications. PLINK's --homozyg-group and --homozyg-match commands were used to find allelically matching ROH that shared at least 95% of physical distance of the smaller ROH.

4.2.8 *Froh* Analysis

ROH burden, or *Froh*, represents the proportion of the autosome in ROHs. *Froh* was derived by summing the total length of autosomal ROHs in an individual and dividing this by the total SNP-mappable autosomal distance (2.77×10^9). A fixed-effect multiple regression model controlling for potential genotypic and phenotypic confounds (see below) was used to examine the effect of *Froh* on IQ. IQ was standardized within each dataset, and dataset included as a covariate in the combined data.

From a genotyping standpoint, *Froh* can be affected by population stratification (e.g., if background levels of homozygosity or autozygosity differ across ethnicities), low quality DNA leading to bad SNP calls, and homozygosity levels that differ reliably across plates, DNA

sources, or samples. To control for potential effects of stratification, we included the first ten principal components generated from an identity-by-state matrix derived from the set of SNPs (~50,000) used to determine genotype ethnicity. To account for the independent calculation of principal components in each dataset, principle components were derived separately for each dataset and regressed on standardized IQ, wherein residual scores were used as the dependent variable in the multiple regression model. We also controlled for the percentage of missing calls and excess homozygosity, both of which have been shown to approximate the quality of SNP calls across samples (Laurie et al., 2010). In particular, we used individual SNP missingness determined from raw SNP data, with the exception of the Lothian, Manchester, and Newcastle datasets, where individual SNP missingness was determined from QC cleaned data. This discrepancy, however, is quite minor as a comparison of missingness rates between raw and cleaned SNP data from the six other datasets was highly similar (overall r^2=0.8, p=0). With respect to excess homozygosity, although it is expected that it will co-vary with *Froh*, previous work has found that *Froh* estimates of autozygosity are generally independent of excess homozygosity when genotyping error is low (Keller et al., 2011). However, when genotyping errors do exist, they typically lead to excess heterozygosity (Laurie et al., 2010) and missed ROHs, thus reducing the power of *Froh* effects as opposed to increasing the false positive rate.

Finally, we also controlled for age and gender, as well as the interaction of age with dataset, as any effects from these variables should be independent from the relationship of *Froh* on IQ. The age by dataset interaction was included because we found a significant difference of the effect on age on IQ between samples of young individuals (GAIN, Brisbane, and CADD samples) and older individuals (Lothian, Manchester, and Newcastle samples; beta=0.14, t(4,356)=2.07, p=.04)**.**

Froh burden was also partitioned in short and long ROH, as well as common and uncommon ROH, as different lengths and frequency reflect the age and population genetic mechanisms explaining the existence of an ROH. In both cases, a median split was used to divide ROHs into 'short/long' groupings or 'common/uncommon' groupings. Due to the variation in SNP density across dataset platforms, median splits for both length and frequency were calculated within each dataset. Median splits for each dataset can be found in Table 4.2. Across all datasets, short ROHs made up 34% of *Froh*, while long ROHs made up 66% of *Froh*. Moreover, common ROHs made up 38% of *Froh*, whereas uncommon ROH made up 62% of *Froh*.

Table 4.2 *Froh* Descriptive Statistics

Dataset	N	Region	Median ROH Length (kb)	% Short ROH	% Long ROH	Median ROH Count	% Common ROH	% Uncommon ROH
IBG	301	Colorado, USA	1076	35.4%	64.6%	1	38.5%	61.5%
GAIN NE	357	Northern Europe	1170	36.4%	63.6%	1	26.4%	73.6%
GAIN UK	183	United Kingdom	1304	39.7%	60.2%	1	9.6%	90.2%
GAIN SP	68	Spain	1280	44.4%	55.1%	1	7.4%	92.1%
LBC 1921	495	Scotland	1032	33.6%	66.4%	1	46.2%	53.8%
LBC 1936	939	Scotland	1009	33.2%	66.8%	2	42.7%	57.3%
MANC	763	England	1018	33.3%	66.6%	2	35.9%	64.1%
NEWC	717	England	1013	33.3%	66.7%	2	34.7%	65.3%
QIMR	537	Australia	998	34.5%	65.5%	1	48.4%	51.9%
Total	4360	---	---	34.3%	65.7%	---	37.8%	62.1%

4.2.9 ROH Mapping Analysis

To see if any particular genomic region predicted IQ, genome-wide ROH mapping procedures were tested using three different segment length thresholds. For each procedure, the autosome was divided into regions of 1 Mb (2,878 tests), 500 kb (5,747 tests), or 250 kb (11,483 tests). Within each genomic region, a linear regression model on IQ was performed on whether or not individuals had an ROH at any given location in that region. IQ in this model was residualized for all covariates in the multiple regression model described above. To derive a genome-wide significance threshold corrected for multiple testing, fully residualized IQ scores were permuted within the combined dataset and regressions run across every region in the genome, preserving the most significant result of each permutation. The permutation process was repeated 1,000 times. The 50^{th} most significant p-value was the genome-wide significance threshold and the 100^{th} most significant p-value was the "suggestive" genome-wide significance threshold. Thus, given the null hypothesis, there was a 5% (10%) chance of observing a single genome-wide significant (suggestive) hit.

4.2.10 Polygenic ROH Mapping

Polygenic ROH mapping is a rank-order approach to test for an excess of negative or positive ROH regions among more significantly predicting regions of the genome. Using regions defined from the ROH mapping approach, regions were rank-ordered according to statistical significance. Starting from the most significant region, each region is added in a sequential manner one-by-one, with the overall count of positive and negative slopes ascertained at each interval. To measure significance, a binomial exact test is run to measure the deviation in the proportion of positive and negative slopes from the overall proportion observed in the genome. For the binomial exact test, both z-scores and corresponding *p*-values were ascertained. To

control for non-independence between ROH regions, IQ scores were permuted 1,000 times and empirical *p*-values were derived. To control for whole-genome *Froh* on rank-order, β estimates were centered and statistical significance was re-calibrated for each region. All *p*-values are two-sided. To test the effect of a reduced genome *Froh* estimate, ten-fold cross validation on the sample used a 90% training sample and 10% test sample. Equal proportions of individuals from each dataset were randomly assigned within a given test set, with leftover individuals from each sample randomly assigned to a test set after all ten test sets were created. For each round of cross-validation, rank-orders of ROH regions were determined by the 90% training set, and *Froh* burden was measured within these regions in the 10% test set. After all ten test sets were trained, the test sets were combined and *Froh* estimates were regressed onto residualized IQ scores (see Table 4.4).

4.3 Results

After quality control procedures within each of the nine datasets, we had genotype and IQ data for 4,360 individuals of European ancestry from five geographic regions: Northern Europe (n=357), United Kingdom (n=3,097), Spain (n=68), Australia (n=537), and the United States (n=301). In particular, the United Kingdom region compromises five separate samples, with one sample using a different genotyping platform from the other four. Four recommended parameters were used in PLINK to call autozygous ROH (see *Methods* for details), with genotype data lightly pruned for linkage disequilibrium (LD) and a minimum of 65 homozygous SNPs showing the strongest relationship to IQ. Under this ROH calling procedure, the mean percentage of the genome in ROHs (*Froh*) was 0.41% (sd=0.36%) in the full sample, with a mean ROH length reaching 1.3 Mb (see Figure 4.1 for full distribution). Within each dataset, the IQ measure was

standardized, then residualized on the first ten principle components generated by the multi-dimensional scaling algorithm in PLINK (Figure 4.2). Tests of genome-wide ROH burden and ROH association on IQ controlled for sex, age, dataset, the inbreeding coefficient (*F*), SNP missingness, an SNP missingness by dataset interaction, and an age by dataset interaction as covariates.

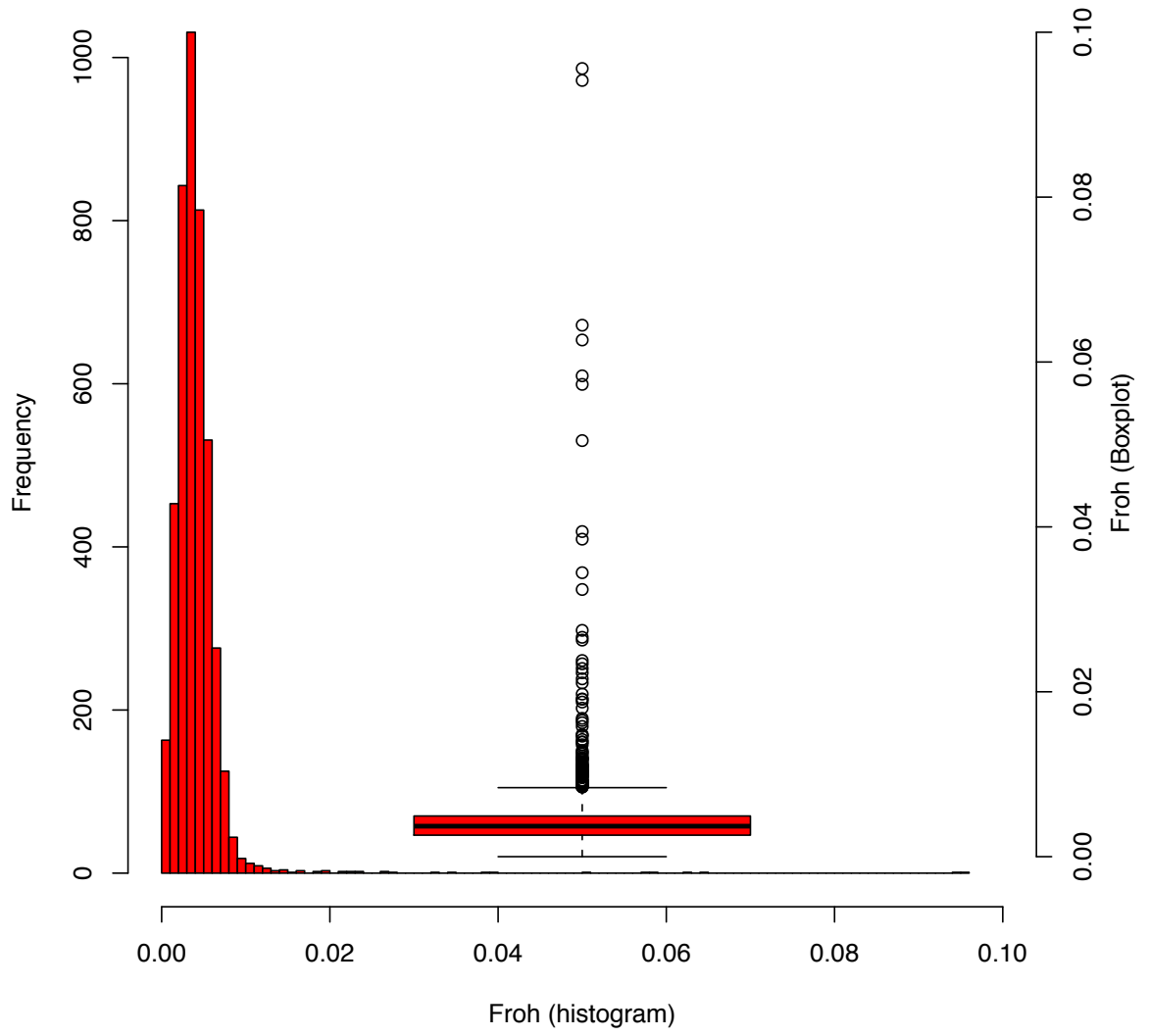

Figure 4.1 Distribution of *Froh* for Combined Sample. The x-axis measures the histogram, and the z-axis measures the boxplot.

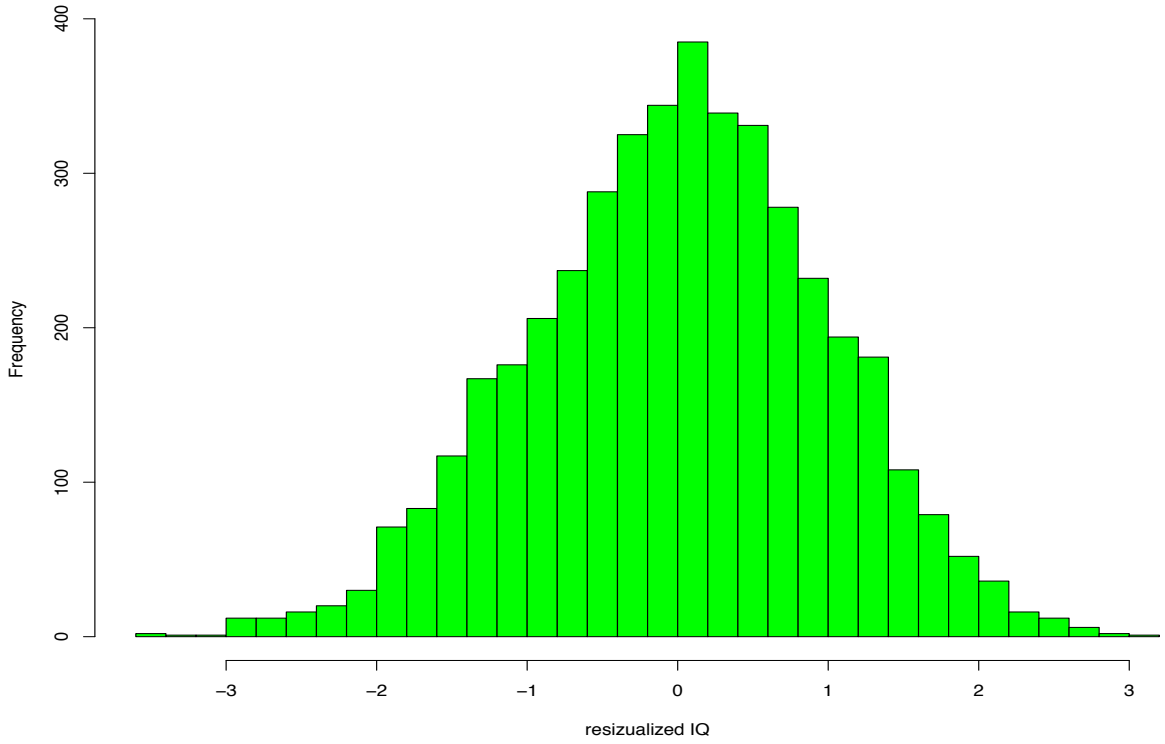

Figure 4.2 Distribution of IQ for Combined Sample. Prior to inclusion in the combined sample, IQ in each dataset is first standardized, then residualized on the first ten principal components from the genotype data.

4.3.1 *Froh* Results

In the combined sample, a fixed effects linear regression model predicting IQ was used, with *Froh* showing a negative, but non-significant relationship to IQ (β = -5.63, $t(4,330) = -1.12$, $p = 0.26$). The point estimate of *Froh* predicts IQ to drop by ~0.84 points for every 1% increase in *Froh*, extrapolating to a drop of ~5.28 IQ points expected on average in the offspring of first cousin inbreeding. Figure 4.3 shows the β and 95% confidence intervals within each dataset as well as the combined sample. Only the GAIN United Kingdom dataset found a significant negative relationship of *Froh* on IQ, whereas none of the datasets found a significant positive relationship. Taken together, eight of the nine slope estimates were negative, which was not expected by chance (exact binomial test, $p = 0.001$). Despite the non-significant result, the relationship of *Froh* on IQ is quite robust, as it was not affected by individuals at the extremities of *Froh*, and changed little after trimming the 25 highest and 25 lowest subjects on *Froh*, (β = -5.16, $t(4,280) = -1.03$, $p = 0.31$), as well as after trimming the 50 highest and 50 lowest subjects on *Froh* (β = -5.4, $t(4,230) = -1.07$, $p = 0.28$). In addition, results above a 50 SNP minimum threshold for calling ROH remained fairly stable, whereas smaller thresholds removed the negative trend (Figure 4.4). Additional analyses removing covariates in step-wise fashion, as well as splitting the sample by age or sex, were not found to moderate the effect of *Froh* on IQ.

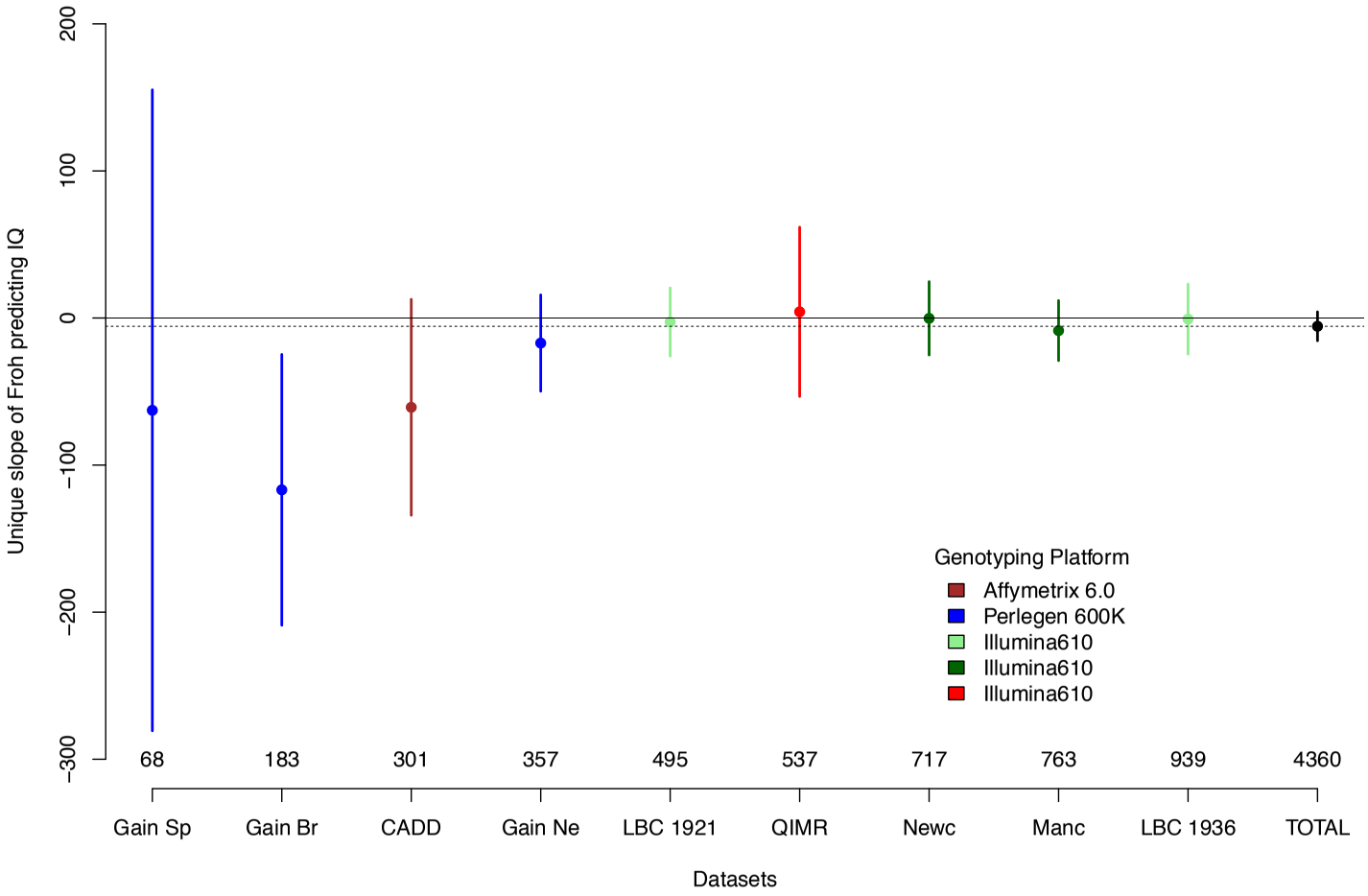

Figure 4.3 *Froh* Association with IQ Across Samples and in the Combined Sample. Points represent the unstandardized beta estimate and lines represent the 95% confidence interval. Dataset size is shown above the x-axis.

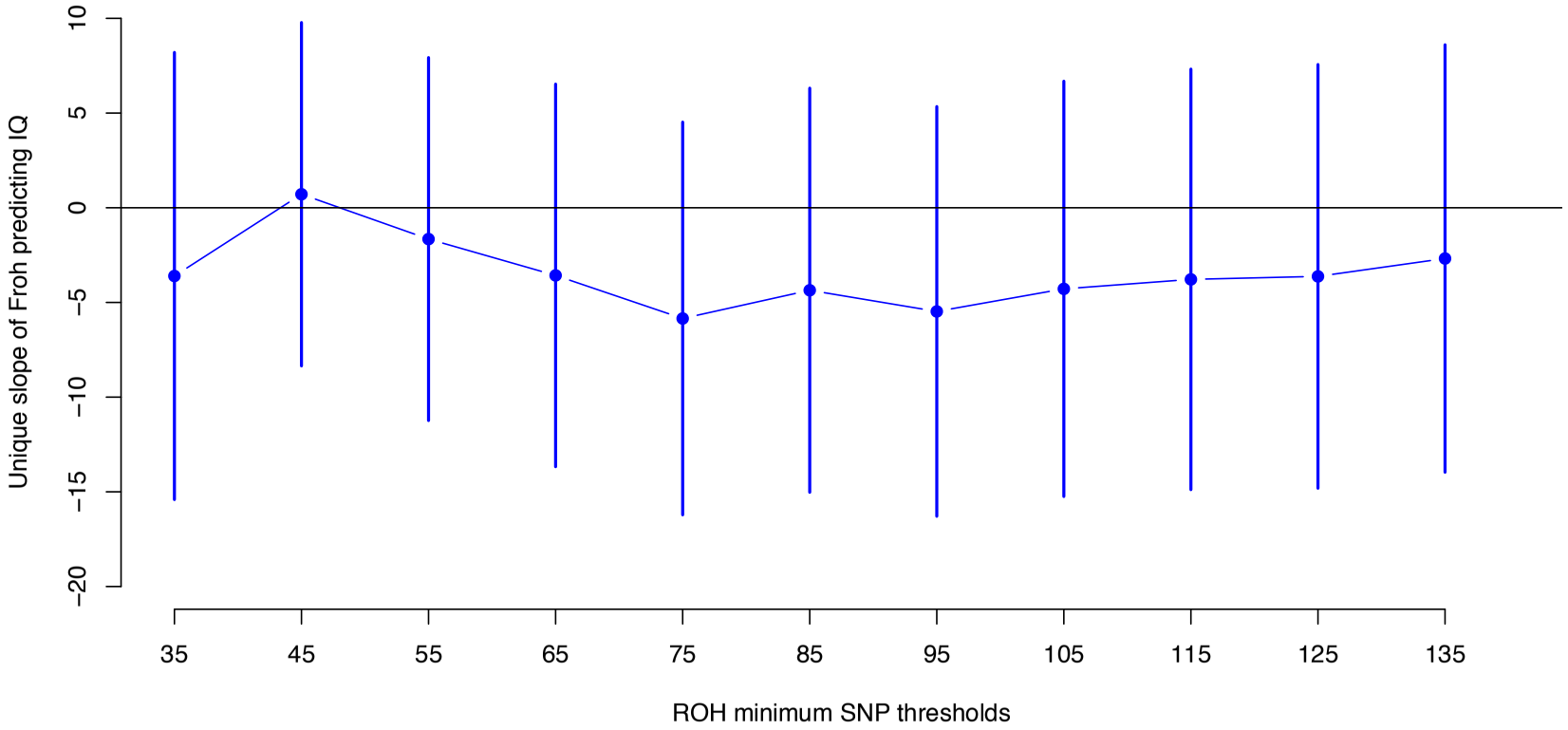

Figure 4.4 *Froh* Association with IQ Across Various Minimum SNP Thresholds. Estimates are only for the combined sample. Points represent the unstandardized beta estimate and lines represent the 95% confidence interval.

In addition to its inclusion as a covariate the main *Froh* analysis, the inbreeding coefficient (*F*) was also examined separately as a predictor of both *Froh* and IQ. As mentioned in the introduction, a positive value for *F* is another signature of inbreeding, albeit at a much more distant level ancestrally and less powerful than *Froh* at measuring inbreeding levels (Keller et al., 2011). In addition, expected multicollinearity between *Froh* and *F* as predictors of IQ may affect the interpretation of the results. As expected, *F* was strongly related to *Froh* when controlling only for dataset ($\beta = 0.09$, $t(4,358) = 22.6$, $p = 0$), however it did not significantly predict IQ in the full model ($\beta = 1.31$, $t(4,330)=0.5$, $p=0.62$), nor did it have much effect when removed from the full model of *Froh* on IQ ($\beta = -4.26$, $t(4,331) = -1.01$, $p = 0.31$). Another possible concern was the chance of hemizygosity due to deletions or partial uniparental disomy being included in the *Froh* estimates. Nevertheless, prior research using information on call intensities found that fewer than 0.3% of the total lengths of ROHs in their samples were actually hemizygous, suggesting that deletions would have a minimal effect on the present results (McQuillan et al., 2008). This finding is not unexpected given the typical size of ROHs in the current sample (in the hundreds or thousands of kb) versus the typical sizes of deletions (in the tens of kb; McCarroll et al., 2005).

Along with overall *Froh* burden on IQ, *Froh* was partitioned by length (long and short ROH) as well as frequency (common and uncommon ROH) using a median split within each dataset (see table 4.3 for details). Examining different ROH lengths provides a direct measure of the relative effects of inbreeding ancestry, as offspring from parents who share a recent common ancestor should carry longer ROHs on average than offspring from parents who share more distant common ancestors. Contrary to expectations, *Froh* arising from long ROH ($\beta = -4.14$, $t(4,330) = -0.82$, $p = .41$) was less predictive of IQ than *Froh* arising from short ROH ($\beta =$ -

38.85, $t(4,330) = -1.53$, $p = .12$). Neither estimate was predictive, however, indicating that neither recent nor distant inbreeding are differentially affecting the *Froh*-IQ relationship. With respect to ROH frequency, common ROH can often represent homozygous haplotypes shared by many unrelated people in each sample—which might indicate positive selection rather than recent inbreeding (Voight, Kudaravalli, Wen, & Pritchard, 2006a), whereas uncommon ROH are more likely to pair up individually rare recessive or partially recessive mutations thought to be scattered throughout the genome (Fay, Wyckoff, & Wu, 2001) within an extended lineage. Consistent with expectations, *Froh* arising from common ROH (beta=-7.44, $t(4,330)=-0.72$, $p=.47$) was less predictive of IQ than *Froh* arising from uncommon ROH (beta=-7.59, $t(4,330)=-1.07$, $p=.28$). As with ROH length, neither estimate was predictive, indicating that ROH frequency does not substantially impact the *Froh*-IQ relationship.

4.3.2 ROH Mapping Results

In contrast to *Froh* burden, ROH mapping can identify distinct genomic regions carrying highly penetrant recessive alleles that affect IQ. Figure 4.5 shows a Manhattan plot of the association *p*-values from ROH mapping using 1 Mb regions, and figure 4.6 shows the QQ plot of the association *p*-values. Among the three size thresholds, no region surpassed the permuted significance threshold (permuted $\alpha > 0.05$) or the suggestive significance threshold (permuted $\alpha > 0.1$). However, among the top 100 regions, those negatively associated with IQ were over-represented in all three analyses. With a 52% base rate overall of negative slopes in the 1 Mb analysis, 64 of the top 100 ROHs were negatively associated to IQ (exact binomial test, $p = 0.01$), suggesting that while no ROH are significant after genome-wide correction, there may be evidence that ROH are exerting slight deleterious effects on IQ at particular regions across the genome. Because it is unlikely that the entire genome is equally contributing to the genetic

variance underlying IQ, the finding above prompted further investigation to see if *Froh* estimates restricted to regions harboring an excess of negative effects would reduce some of the statistical noise damping down the whole-genome *Froh* effects.

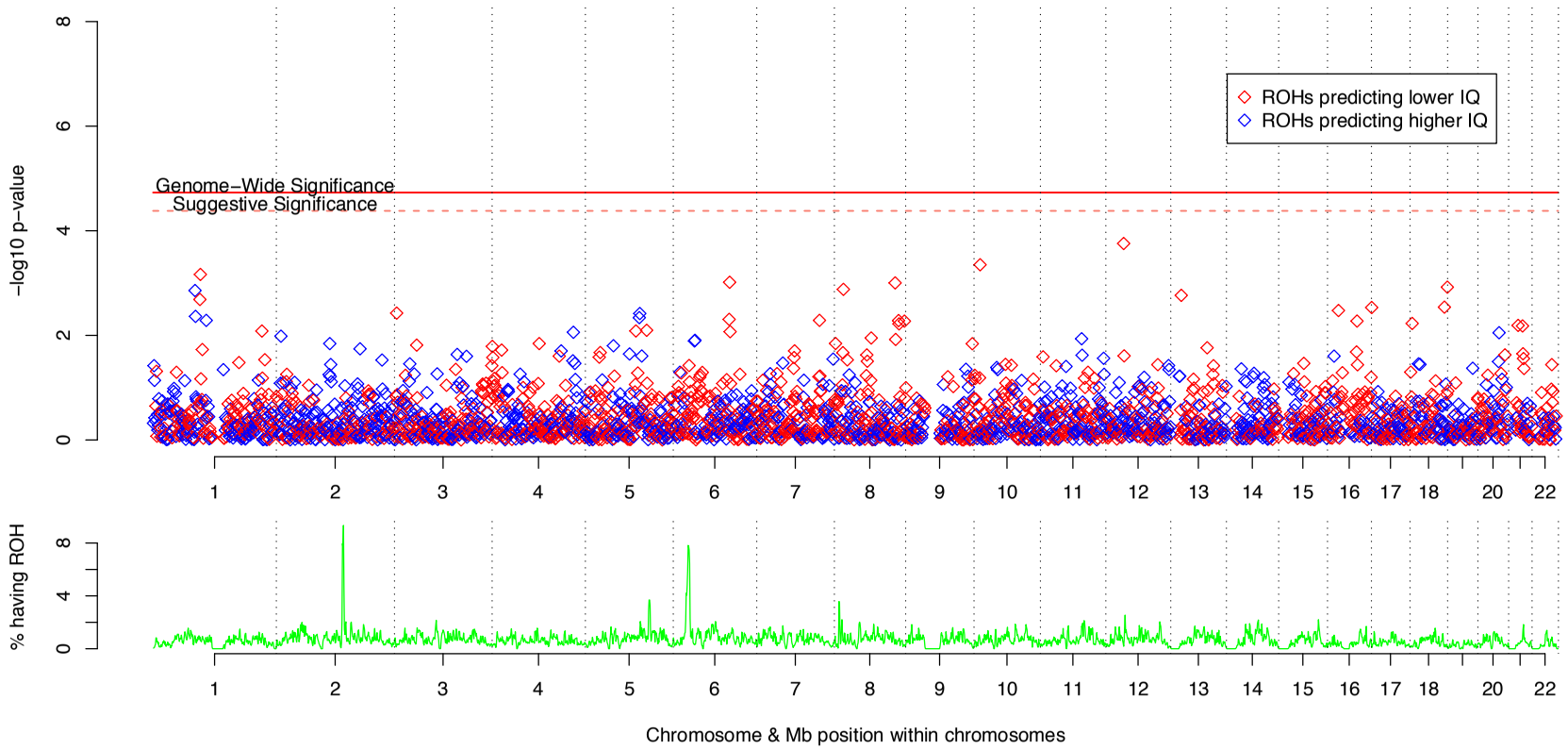

Figure 4.5 Manhattan Plot of 1Mb ROH Regions Predicting IQ**.** Top panel: -$\log_{10}$ *p*-values for 1Mb ROH regions predicting lower (red) and higher IQ (blue). Each region controls for all covariates prior to association analysis. Bottom panel: ROH frequencies for each region across the autosome.

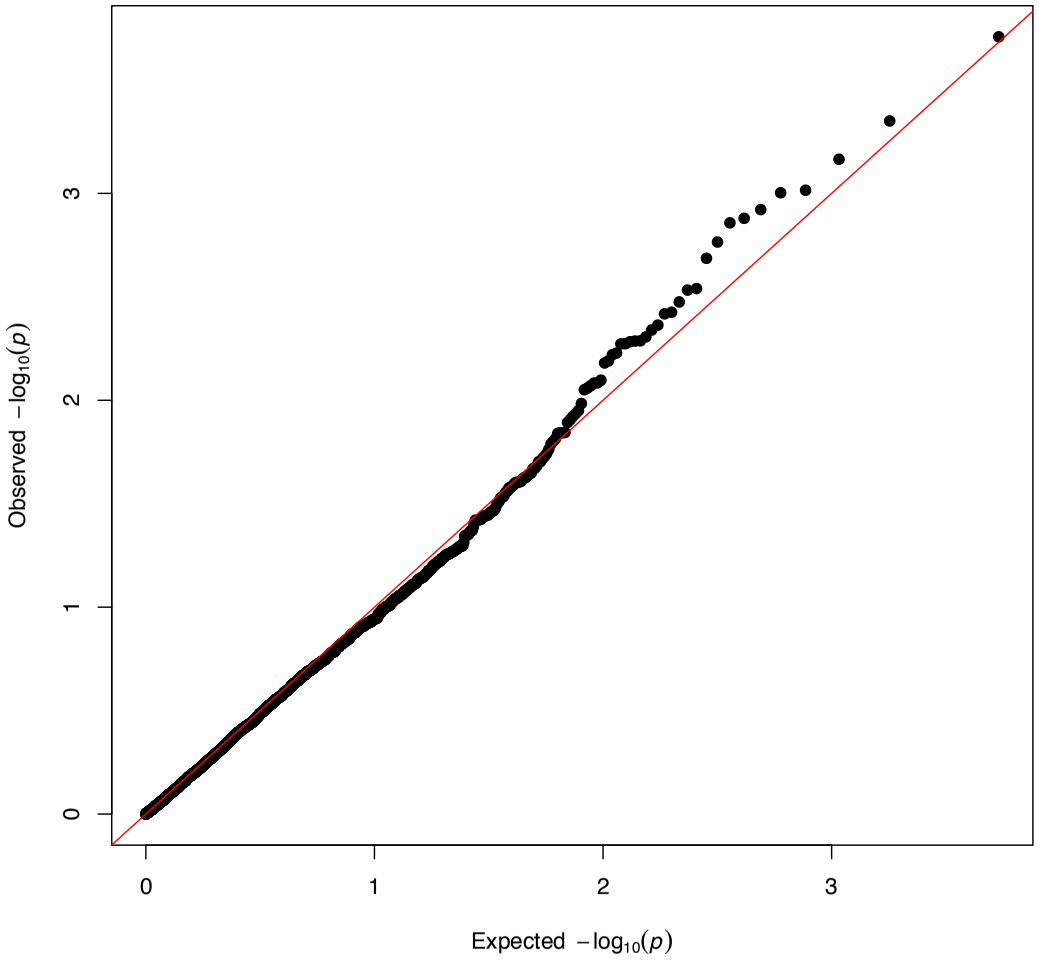

Figure 4.6 QQ Plot of 1Mb ROH Regions Predicting IQ**.** Each region controls for all covariates prior to association analysis.

4.3.3 Polygenic ROH Mapping

The approach, termed here as polygenic ROH mapping, begins by ranking 1 Mb ROH map regions by statistical significance. Starting from the most significant region, each region is added in a sequential manner one-by-one (e.g. top 10 regions, top 11 regions, etc.), with the overall count of positive and negative slopes ascertained at each interval. To measure significance, a binomial exact test is run to measure the deviation in the proportion of positive and negative slopes from the overall proportion observed in the genome (see *Methods* for details). Nearby regions can be affected by a single long ROH and produce non-independent signals, so empirical *p*-values were derived through permutation of IQ scores. Whole-genome *Froh* effects can also substantially skew the relative proportions of more significant regions, so re-calibrated significance of each region controlling for *Froh* burden are also reported. However, because the *Froh* burden effect in this study is small, the rank-order of regions is very similar between original and re-calibrated significance ($r = 0.98$, $p = 0$). Figure 4.7 shows the polygenic ROH mapping results using 1 Mb regions, with z-scores representing the direction of the effect towards an excess of positive or negative slopes. Both rankings show a significant excess of ROH regions predicting lower IQ (i.e. negative slopes and negative z-score) among the top regions, indicating that the effect of ROH on IQ is restricted to a particular subset of more significant regions of the genome. Among the top regions, two significant spikes of ROH regions predicting lower IQ emerge; the top 31 regions (permuted $p = 0.003$) and the top 434 regions (permuted $p = 0.01$). Similar spikes are found with the re-calibrated significance ranking; the top 24 regions (permuted $p = 0.012$) and the top 415 regions (permuted $p = 0.051$).

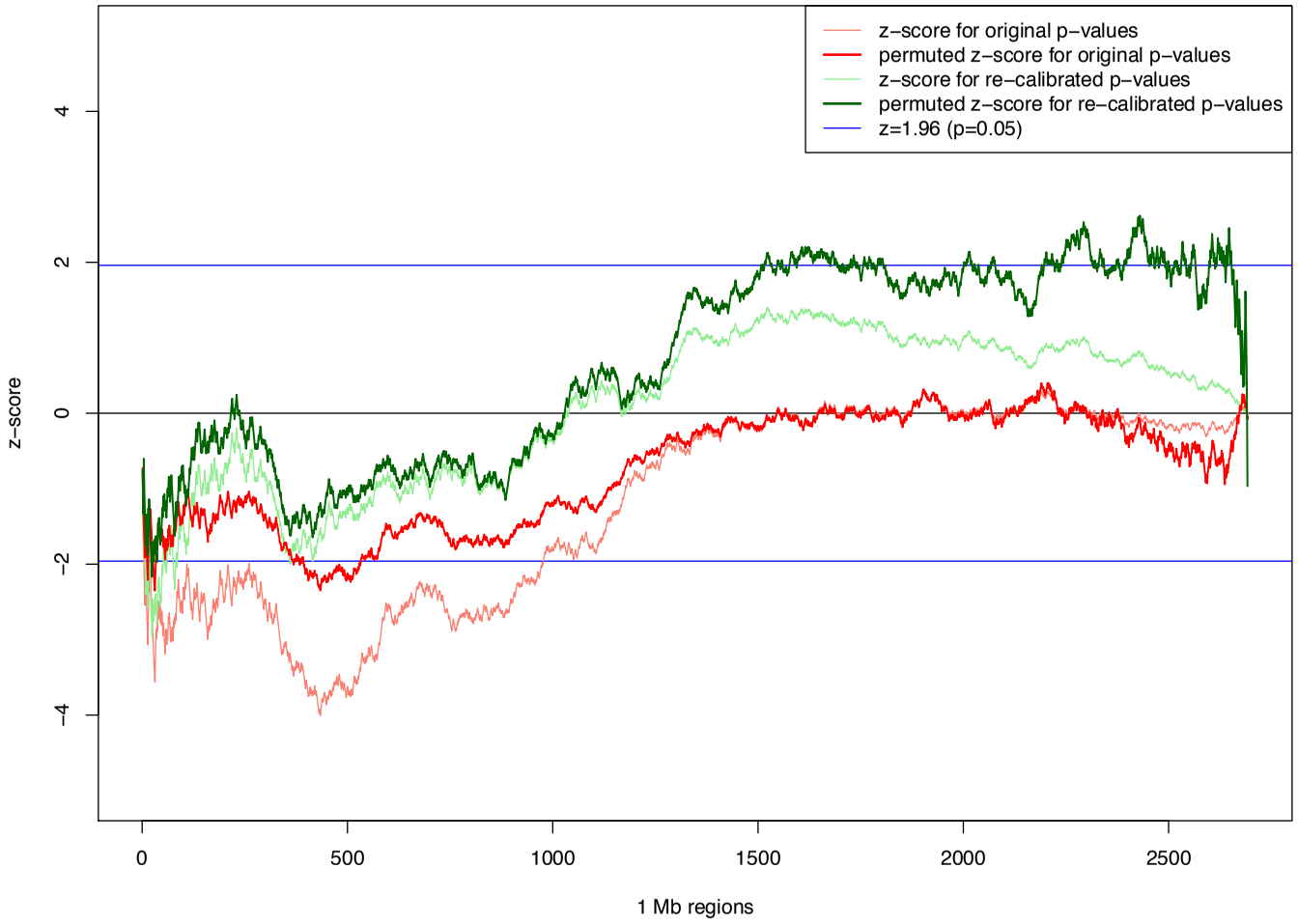

Figure 4.7 Binomial Test for Excess of ROH Predicting Higher IQ (z-score > 0) or Lower IQ (z-score < 0) Among Rank-Ordered 1 Mb Regions. Both original rank-order *p*-values (red) and re-calibrated rank-order *p*-value rankings (green) are shown.

In order to test the how well the recommended regions above predict IQ, ten-fold cross validation was implemented to generate an unbiased estimate of variance explained within the current sample. The hypothesis is that the recommended subset of ROH regions will predict IQ better than whole-genome *Froh* burden. The 90% training set was used to generate the rank-order of ROH regions, and *Froh* burden was calculated in the 10% test set among recommended subsets of regions. The top 1,000 and 2,000 ROH regions were also included to compare against the recommended subsets. Table 4.4 shows the variance explained by *Froh* burden at various ROH region subsets, with the top 1,000 regions rank-ordered by the original *p*-values showing the highest R^2 estimate ($R^2 = 0.0019$, $p = 0.004$). Overall, the recommended estimates did perform better than whole-genome *Froh*, but did not out-perform the top 1,000-region comparison, suggesting that peak significance using an exact binomial test are only a rough guide to determining the best practices to reduce the search space for maximizing ROH effects in IQ. In general, while the variance explained for IQ using the polygenic ROH mapping approach is still quite small at 0.19%, it is a 6.8-fold increase from whole-genome *Froh* (0.028%), providing evidence that whole-genome estimates include a substantial amount of ROH variance that has no effect on IQ. The rank-ordered ROH region sets were also examined to see if there was an enrichment of genes expressed in the central nervous system (CNS genes). The subset used gathered genes using four separate databases that show either differential CNS expression, known neuronal activity, and/or roles in learning and synapse function, and compromise roughly 5% of known genes (Allen et al., 2010; Lee et al., 2012). Chi-squared tests of each region set, however, showed no evidence of CNS gene enrichment relative to the full genome (all $p > 0.05$).

Table 4.4 Cross-Validation Results of *Froh* Restricted to Top Regions

Subset of ROH regions	**β**	***p*-value**	R^2
Rank-ordered by original p-values			
Top 31 regions	-0.98	0.520	9.50e-5
Top 434 regions	-9.57	0.012	0.0015
Top 1,000 regions	-13.13	0.004	0.0019
Top 2,000 regions	-8.46	0.090	6.58e-4
Rank-ordered by re-calibrated p-values			
Top 24 regions	-1.60	0.247	3.07e-4
Top 415 regions	-10.03	0.007	0.0016
Top 1,000 regions	-12.86	0.005	0.0018
Top 2,000 regions	-7.92	0.112	5.80e-4

4.4 Discussion

The present study tested the relationship of autozygosity to IQ using high-density SNP data from nine separate sites compromising 4,360 individuals. After stringent quality control and the application of preferred methods for detecting autozygosity, the findings do not support a significant role of autozyosity burden (*Froh*) in explaining the variation in IQ among outbred populations of European ancestry, nor point to any specific locus whereby autozygosity associates with IQ. However, post-hoc analysis of *Froh* estimates restricted to genomic regions more likely to affect IQ show an increase in prediction (from 0.028% to 0.19% in variance explained), as evidenced through cross-validation. The overall direction of *Froh* estimates - with increased *Froh* correlating with lower IQ - is consistent with the prediction that inbreeding depression is likely to have a negative effect cognitive ability. In relation to previous studies of inbreeding on IQ, the estimates observed correspond fairly well with estimates of first cousin inbreeding. In the current study, inbreeding among first cousins leads to an average *Froh* of 6.25%, corresponding to a predicted drop of 5.28 IQ points. With respect to previous research, the effect is roughly one half (Afzal, 1988) to two (Morton, 1978) times the magnitude of known inbreeding lowering IQ. Despite the non-significant overall findings, the negative direction of results does have plausible ramifications for populations where consanguineous marriage is commonplace (Bittles & Neel, 1994).

The findings shed some light on the effect of non-additive genetic variance underlying IQ. Autozygosity by definition captures only non-additive sources genetic variance, although it is constrained to the proportion of the genome that is autozygous. Nonetheless, relative proportions of phenotypic variance explained by autozygosity give a rough measure of non-additive genetic underpinnings. For example, another recent large-scale study of autozygosity on Schizophrenia

(Keller et al., 2012) found a significant, albeit modest effect of increased *Froh* burden predicting Schizophrenia (Nagelkerke's R^2 = .00068), while the current effect size estimate for *Froh* burden predicting IQ is even lower (R^2 = .00028). When set side by side with recent measures of additive genetic variance, the findings above correlate to some extent with the reported proportions of heritability explained by SNPs for Schizophrenia (23% explained by SNPs; Lee et al., 2012) and for IQ (51% explained by SNPs; Davies et al., 2011). While these comparisons by no means represent a definitive means of partitioning relative additive and non-additive effects on genotypic variance, they reaffirm recent findings that a large extent of genotypic variance in IQ is additive, and non-additive measures, such as autozygosity burden, are restricted in part to a smaller proportion of the remaining heritability to be explained.

4.4.1 Limitations

There are certain limitations in the study that must be addressed. Prior to the acquisition of the current dataset, it was known that ostensibly outbred populations required very large sample sizes, on the order of 20,000 to 60,000, to obtain adequate power to detect the effects of inbreeding depression (Keller et al., 2011). Despite the current sample falling well below these numbers, it is still the largest consortium of data to date measuring IQ and autozygosity. Nevertheless, individuals from outbred populations generally exhibit low variation in overall autozygosity, so larger samples would be required to affirm if the signal in the current dataset is a true effect. An alternative to increasing sample size would be to collect data from samples targeted towards populations with higher variance in inbreeding, such as regions where consanguineous marriage rates are higher, or by ascertaining samples that are on the extreme end of IQ, such as individuals with intellectual disability. Statistical noise and possible bias represents another limitation of the study, as our operational construct of IQ is stitched together

from different cognitive tests and measures standardized across each dataset. While IQ measures are well known for their robust convergent validity (Johnson et al., 2008), there will still be a level of uncertainty that can add noise to the data and decrease statistical power. With regards to the genotype collection and SNP calling, while strict QC procedures were followed to minimize genotyping error, the use of different platforms across datasets nonetheless introduced additional noise and possible biases that are unable to be controlled. For example, overall autozygosity measurements from the three datasets making up the GAIN IMAGE project are systematically lower than the other datasets, as the earlier genotyping platform had roughly one half to quarter the SNPs to begin with than the other genotyping chips.

4.4.2 Conclusion

Autozygosity is the central measure of inbreeding depression, and can help researchers understand the both the genetic architecture and population genetic history underlying well-known heritable traits such as IQ. The current study represents the first to measure IQ and autozygosity using dense genome-wide SNP data, using an extensive consortia panel from around the world. In summary, whole-genome autozygosity burden and genome-wide autozygosity mapping are not significant predictors of IQ in outbred populations of European descent. However, the relationship found is similar to previous estimates of autozygosity on IQ among consanguineous families, and there is a nominally significant negative relationship of autozygosity burden on IQ when restricted to a subset of the genome rank-ordered by statistical significance. These findings shed light on the relative contribution of inbreeding within the normal spectrum of variation in IQ, and suggest that future investigations target either populations known to have higher inbreeding coefficients or more extreme samples with regards to IQ.

Chapter 5

Summary and Conclusion

5.1 Background

The field of human molecular genetics has undergone a substantial technological transformation in the past decade, allowing researchers to identify and analyze genetic variation across the human genome with unprecedented depth and precision. A central goal in utilizing this technological advancement is to discover the genetic variation that underlies complex heritable traits and disorders. In recent years, much of the focus has been on large-scale genome-wide association studies (GWAS) in an attempt to identify effects of common single nucleotide polymorphisms (SNPs) on a phenotype. Progress in this arena, however, has been limited, as validated findings for most phenotypes represent only a small fraction of the variance attributed to genetics known from family and twin studies, leaving a large proportion of heritability to be explained. My research is in large part motivated by the issues surrounding GWAS, as additional methodological techniques and population genetic theory can help explain phenotypic variance unaccounted for by the traditional genetic association design.

5.2 Approaches to Interrogating Genome-Wide SNP Data

In my first study, I look at a case/control sample of bipolar disorder, examining how prior information from linkage studies can inform GWAS signals. The primary aim is to ease the burden of multiple-testing correction applied in GWAS using empirically informed weighting to tease out true signals supported by prior genetic evidence. In my second study, I determine the best practices to detect signatures of distant inbreeding, via runs of homozygosity, in genome-

wide SNP data. The motivation for studying this phenomenon is due to the extensive evidence of inbreeding depression on fitness related traits, where the effects of recessive or partially recessive alleles are expressed. Using an extensive simulation design, I test multiple programs to determine the optimal method to identify runs of homozygosity caused by distant inbreeding. In my third and final study, I apply my work from the second study to a comprehensive dataset of cognitive measures to understand the extent to which distant inbreeding affects variation in cognitive ability.

5.3 Summary of Results

In my first study, the incorporation of prior linkage information to weight Bipolar Disorder GWAS *p*-values did not result in any significant SNPs after genome-wide correction for multiple testing. However, the top SNP signal from the initial GWAS was further supported as it occurred within a suggestive linkage peak on chromosome 18. In addition, I reported SNPs in the lowest 99^{th} percentile of *p*-values also occurring within the top 95^{th} percentile of linkage *z*-scores as regions of interest.

My second study, which investigated the accuracy of autozygosity detection for three separate programs, recommended PLINK over GERMLINE and BEAGLE when detecting autozygous segments. Furthermore, SNP data pruned for high levels of linkage disequilibrium (LD) also increased the accuracy of detection. The results provided both general and specific recommendations for maximizing autozygosity detection in genome-wide SNP data, and should apply equally well to research on whole-genome autozygosity burden as well as autozygosity in targeted regions.

In my third study, I applied the recommendations from my second study to large-scale consortia SNP data to test the relationship of autozygosity to general cognitive ability (IQ).

Overall, the findings do not support a significant role of autozyosity burden (*Froh*) in explaining the variation in IQ among outbred populations, nor point to any specific locus whereby autozygosity associates with IQ. However, the negative direction and effect size of the relationship does fall in line with earlier measures of consanguineous inbreeding lowering IQ. In addition, cross-validation of *Froh* estimates restricted to genomic regions more likely to affect IQ show up to a 7-fold increase in prediction, suggesting that the effects of autozygosity on IQ are not uniform across the genome.

5.4 Conclusion

In general, genome-wide SNP association has been relatively unsuccessful in uncovering the genetic etiology of traits such as Bipolar Disorder and IQ, yielding only scant results that fail to explain even a moderate proportion of trait variance. In turn, this difficulty has prompted alternative approaches to uncovering the missing heritability. Overall, my application of alternative methods of SNP interrogation, including the inclusion of prior linkage information and effects of autozygosity in outbred populations, did not come away with any substantial increase in explaining trait variation. While these results reflect the complexity of psychiatric phenotypes and difficulty in teasing out their genetic architecture, recent successes in this arena have come from approaches that investigate the degree to which a trait is under highly polygenic genetic architecture.

One approach, which looked at both Schizophrenia and Bipolar disorder, showed that predictive ability significantly increased when validating the top percentiles of SNP associations in a large case-control Schizophrenia dataset to other datasets in both Schizophrenia and Bipolar Disorder (Purcell et al., 2009). More recently, a GWAS on IQ has found that quantifying the variance accounted for using all SNPs captures a substantial proportion of heritable variation

(Davies et al., 2011), suggesting that IQ has a highly polygenic architecture. While the discovery of the genetic variants for Bipolar Disorder and IQ has so far been a daunting task, emerging technologies such as whole-genome sequence data along with recent successes of estimating polygenic contributions to trait variance show promise in finally uncovering the molecular basis underlying heritable variation in these traits.

References

Consortium, G. P. (2010). A map of human genome variation from population-scale sequencing. *Nature*, *467*(7319), 1061-1073.

Abecasis, G. R., Cherny, S. S., Cookson, W. O., & Cardon, L. R. (2002). Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet*, *30*(1), 97-101.

Abu Safieh, L., Aldahmesh, M. A., Shamseldin, H., Hashem, M., Shaheen, R., Alkuraya, H. et al. (2010). Clinical and molecular characterisation of Bardet-Biedl syndrome in consanguineous populations: the power of homozygosity mapping. *J Med Genet*, *47*(4), 236-241.

Afzal, M. (1988). Consequences of consanguinity on cognitive behavior. *Behavior genetics*, *18*(5), 583-594.

Afzal, M. (1993). Inbreeding depression and intelligence quotient among North Indian children. *Behavior genetics*, *23*(4), 343-347.

Agrawal, N., Sinha, S. N., & Jensen, A. R. (1984). Effects of inbreeding on Raven matrices. *Behavior Genetics*, *14*(6), 579-585.

Allen, H. L., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F. et al. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, *467*(7317), 832-838.

Altmüller, J., Palmer, L. J., Fischer, G., Scherb, H., & Wjst, M. (2001). Genomewide scans of complex human diseases: true linkage is hard to find. *The American Journal of Human Genetics*, *69*(5), 936-950.

Arning, L., Monte, D., Hansen, W., Wieczorek, S., Jagiello, P., Akkad, D. A. et al. (2008). ASK1 and MAP2K6 as modifiers of age at onset in Huntington's disease. *J Mol Med*, *86*(4), 485-490.

Balding, D. J. (2006). A tutorial on statistical methods for population association studies. *Nature Reviews Genetics*, *7*(10), 781-791.

Barrett, J. C., Hansoul, S., Nicolae, D. L., Cho, J. H., Duerr, R. H., Rioux, J. D. et al. (2008). Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nature genetics*, *40*(8), 955-962.

Bashir, R., Britton, S., Strachan, T., Keers, S., Vafiadaki, E., Lako, M. et al. (1998). A gene related to Caenorhabditis elegans spermatogenesis factor fer-1 is mutated in limb-girdle muscular dystrophy type 2B. *Nat Genet*, *20*(1), 37-42.

Batty, G. D., Deary, I. J., & Gottfredson, L. S. (2007). Premorbid (early life) IQ and later mortality risk: systematic review. *Annals of epidemiology*, *17*(4), 278-288.

Baum, A. E., Akula, N., Cabanero, M., Cardona, I., Corona, W., Klemens, B. et al. (2008). A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol Psychiatry*, *13*(2), 197-207.

Bittles, A. H., & Neel, J. V. (1994). The costs of human inbreeding and their implications for variations at the DNA level. *Nature genetics*, *8*(2), 117-121.

Borck, G., Ur Rehman, A., Lee, K., Pogoda, H. M., Kakar, N., von Ameln, S. et al. (2011). Loss-of-function mutations of ILDR1 cause autosomal-recessive hearing impairment DFNB42. *Am J Hum Genet*, *88*(2), 127-137.

Bouchard, T. J., & McGue, M. (1981). Familial studies of intelligence: A review. *Science*, *212*(4498), 1055.

Browning, S. R., & Browning, B. L. (2010a). High-resolution detection of identity by descent in unrelated individuals. *The American Journal of Human Genetics*, *86*(4), 526-539.

Browning, S. R., & Browning, B. L. (2010b). High-resolution detection of identity by descent in unrelated individuals. *Am J Hum Genet*, *86*(4), 526-539.

Butcher, L. M., Davis, O. S. P., Craig, I. W., & Plomin, R. (2008). Genome-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500K single nucleotide polymorphism microarrays. *Genes, Brain and Behavior*, *7*(4), 435-446.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N. et al. (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature genetics*, *22*(3), 231-238.

Carothers, A. D., Rudan, I., Kolcic, I., Polasek, O., Hayward, C., Wright, A. F. et al. (2006). Estimating human inbreeding coefficients: comparison of genealogical and marker heterozygosity approaches. *Annals of human genetics*, *70*(5), 666-676.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge Univ Pr.

Chabris, C. F., Hebert, B. M., Benjamin, D. J., & Beauchamp…, J. Most Reported Genetic Associations with General Intelligence Are Probably False Positives. *economics.harvard.edu*.

Chadeau-Hyam, M., Hoggart, C. J., O'Reilly, P. F., Whittaker, J. C., De Iorio, M., & Balding, D. J. (2008). Fregene: simulation of realistic sequence-level data in populations and ascertained samples. *BMC Bioinformatics*, *9*, 364.

Charlesworth, D., & Willis, J. H. (2009). The genetics of inbreeding depression. *Nat Rev Genet*, *10*(11), 783-796.

Cho, H. J., Meira-Lima, I., Cordeiro, Q., Michelon, L., Sham, P., Vallada, H. et al. (2005). Population-based and family-based studies on the serotonin transporter gene polymorphisms and bipolar disorder: a systematic review and meta-analysis. *Mol Psychiatry*, *10*(8), 771-781.

Cirulli, E. T., Kasperaviči&umacr, D., Attix, D. K., Need, A. C., Ge, D., & Greg Gibson, D. B. G. (2010). Common genetic variation and performance on standardized cognitive tests. *European Journal of Human Genetics*, *18*(7), 815-820.

Collin, R. W., Safieh, C., Littink, K. W., Shalev, S. A., Garzozi, H. J., Rizel, L. et al. (2010). Mutations in C2ORF71 cause autosomal-recessive retinitis pigmentosa. *Am J Hum Genet*, *86*(5), 783-788.

Crum, R. M., Anthony, J. C., Bassett, S. S., & Folstein, M. F. (1993). Population-based norms for the Mini-Mental State Examination by age and educational level. *JAMA: the journal of the American Medical Association*, *269*(18), 2386-2391.

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D. et al. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry*, *16*(10), 996-1005.

Davis, O. S. P., Butcher, L. M., Docherty, S. J., Meaburn, E. L., Curtis, C. J. C., Simpson, M. A. et al. (2010). A three-stage genome-wide association study of general cognitive ability: hunting the small effects. *Behavior genetics*, *40*(6), 759-767.

Deary, I. J. (2012). Human Intelligence. *Annual Review of Psychology*, *63*(1).

Deary, I. J., Johnson, W., & Houlihan, L. M. (2009). Genetic foundations of human intelligence. *Human Genetics*, *126*(1), 215-232.

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. *Journal of Personality and Social Psychology; Journal of Personality and Social Psychology*, *86*(1), 130.

Deary, I. J., Yang, J., Davies, G., Harris, S. E., Tenesa, A., Liewald, D. et al. (2012). Genetic contributions to stability and change in intelligence from childhood to old age. *Nature*.

DeRose, M. A., & Roff, D. A. (1999a). A comparison of inbreeding depression in life-history and morphological traits in animals. *Evolution,* 1288-1292.

DeRose, M. A., & Roff, D. A. (1999b). A comparison of inbreeding depression in life-history and morphological traits in animals. *Evolution,* 1288-1292.

Detera-Wadleigh, S. D., & McMahon, F. J. (2006). G72/G30 in schizophrenia and bipolar disorder: review and meta-analysis. *Biol Psychiatry*, *60*(2), 106-114.

Dickson, S. P., Wang, K., Krantz, I., Hakonarson, H., & Goldstein, D. B. (2010). Rare variants create synthetic genome-wide associations. *PLoS Biol*, *8*(1), e1000294.

DiLella, A. G., Marvi, J., Brayton, K., & Woo, S. L. C. (1987). An ammo-acid substitution involved in phenylketonuria is in linkage disequilibrium with DNA haplotype 2.

Durbin, R. M., Abecasis, G. R., Altshuler, D. L., Auton, A., Brooks, L. D., Durbin, R. M. et al. (2010). A map of human genome variation from population-scale sequencing. *Nature*, *467*(7319), 1061-1073.

Edvardsen, J., Torgersen, S., Roysamb, E., Lygren, S., Skre, I., Onstad, S. et al. (2008). Heritability of bipolar spectrum disorders. Unity or heterogeneity? *J Affect Disord*, *106*(3), 229-240.

Enciso-Mora, V., Hosking, F. J., & Houlston, R. S. (2010). Risk of breast and prostate cancer is not associated with increased homozygosity in outbred populations. *European Journal of Human Genetics*, *18*, 909-914.

Erickson, R. P., Larson-Thome, K., Valenzuela, R. K., Whitaker, S. E., & Shub, M. D. (2008). Navajo microvillous inclusion disease is due to a mutation in MYO5B. *Am J Med Genet A*, *146A*(24), 3117-3119.

Erlenmeyer-Kimling, L., & Jarvik, L. F. (1963). Genetics and intelligence: A review. *Science*, *142*(3598), 1477.

Fan, J., Ionita-Laza, I., McQueen, M. B., Devlin, B., Purcell, S., Faraone, S. V. et al. (2010). Linkage disequilibrium mapping of the chromosome 6q21-22.31 bipolar I disorder susceptibility locus. *Am J Med Genet B Neuropsychiatr Genet*, *153B*(1), 29-37.

Fay, J. C., Wyckoff, G. J., & Wu, C. I. (2001). Positive and negative selection on the human genome. *Genetics*, *158*(3), 1227-1234.

Fenner, J. N. (2005). Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am J Phys Anthropol*, *128*(2), 415-423.

Ferreira, M. A. R., O'Donovan, M. C., Meng, Y. A., Jones, I. R., Ruderfer, D. M., Jones, L. et al. (2008). Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nature genetics*, *40*(9), 1056-1058.

Fisher, R. A. (1930). *The genetical theory of natural selection.* Clarendon Press, Oxford, UK.

Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A. et al. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, *449*(7164), 851-861.

Gibbs, R. A., Belmont, J. W., Hardenbol, P., Willis, T. D., Yu, F., Yang, H. et al. (2003). The international HapMap project. *Nature*, *426*(6968), 789-796.

Gibson, J., Morton, N. E., & Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. *Hum Mol Genet*, *15*(5), 789-795.

Gignac, G. E. (2006). A Confirmatory Examination of the Factor Structure of the Multidimensional Aptitude Battery. *Educational and psychological measurement*, *66*(1), 136-145.

Gilbody, S., Lewis, S., & Lightfoot, T. (2007). Methylenetetrahydrofolate reductase (MTHFR) genetic polymorphisms and psychiatric disorders: a HuGE review. *Am J Epidemiol*, *165*(1), 1-13.

Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E. et al. (1983). A polymorphic DNA marker genetically linked to Huntington's disease. *Nature*, *306*(5940), 234-238.

Gusev, A., Lowe, J. K., Stoffel, M., Daly, M. J., Altshuler, D., Breslow, J. L. et al. (2009). Whole population, genome-wide mapping of hidden relatedness. *Genome Res*, *19*(2), 318-326.

Hardy, J., & Singleton, A. (2009). Genomewide association studies and human disease. *New England Journal of Medicine*, *360*(17), 1759-1768.

Harpending, H. C., Batzer, M. A., Gurven, M., Jorde, L. B., Rogers, A. R., & Sherry, S. T. (1998). Genetic traces of ancient demography. *Proc Natl Acad Sci U S A*, *95*(4), 1961-1967.

Harris, S., Fox, H., Wright, A., Hayward, C., Starr, J., Whalley, L. et al. (2007). A genetic association analysis of cognitive ability and cognitive ageing using 325 markers for 109 genes associated with oxidative stress or cognition. *BMC genetics*, *8*(1), 43.

Hedrick, P. W. (2011). *Genetics of populations.* Sudbury, MA: Jones & Bartlett Publishers.

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S. et al. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*, *106*(23), 9362.

Hirschhorn, J. N., Lohmueller, K., Byrne, E., & Hirschhorn, K. (2002). A comprehensive review of genetic association studies. *Genetics in Medicine*, *4*(2), 45.

Hopper, J. L. (2002). The Australian twin registry. *Twin Research*, *5*(5), 329-336.

Hosking, F. J., Papaemmanuil, E., Sheridan, E., Kinsey, S. E., Lightfoot, T., Roman, E. et al. (2010). Genome-wide homozygosity signatures and childhood acute lymphoblastic leukemia risk. *Blood*, *115*(22), 4472.

Howrigan, D., Simonson, M., & Keller, M. (2011). Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. *BMC genomics*, *12*(1), 460.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med*, *2*(8), e124.

Jackson, D. N. (1998). Multidimensional Aptitude Battery-II (MAB-II).

Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence*, *32*(1), 95-107.

Johnson, W., Nijenhuis, J., & Bouchard, T. J. (2008). Still just 1< i> g: Consistent results from five test batteries. *Intelligence*, *36*(1), 81-95.

Kalay, E., Yigit, G., Aslan, Y., Brown, K. E., Pohl, E., Bicknell, L. S. et al. (2011). CEP152 is a genome maintenance protein disrupted in Seckel syndrome. *Nat Genet*, *43*(1), 23-26.

Keller, M. C., Visscher, P. M., & Goddard, M. E. (2011). Quantification of Inbreeding Due to Distant Ancestors and Its Detection Using Dense Single Nucleotide Polymorphism Data. *Genetics*, *189*(1), 237-249.

Keller, M. C., Simonson, M. A., Ripke, S., & Neale…, B. M. (2012). Runs of Homozygosity Implicate Autozygosity as a Schizophrenia Risk Factor. *PLoS Genetics*.

Kerem, B., Rommens, J. M., Buchanan, J. A., Markiewicz, D., Cox, T. K., Chakravarti, A. et al. (1989). Identification of the cystic fibrosis gene: genetic analysis. *Science*, *245*(4922), 1073.

Kirin, M., McQuillan, R., Franklin, C. S., Campbell, H., McKeigue, P. M., & Wilson, J. F. (2010). Genomic runs of homozygosity record population history and consanguinity. *PLoS One*, *5*(11), e13996.

Kong, A., & Cox, N. J. (1997). Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet*, *61*(5), 1179-1188.

Kremeyer, B., Herzberg, I., Garcia, J., Kerr, E., Duque, C., Parra, V. et al. (2006). Transmission distortion of BDNF variants to bipolar disorder type I patients from a South American population isolate. *Am J Med Genet B Neuropsychiatr Genet*, *141B*(5), 435-439.

Ku, C. S., Naidoo, N., Teo, S. M., & Pawitan, Y. (2011). Regions of homozygosity and their impact on complex diseases and traits. *Human genetics*, *129*(1), 1-15.

Ku, C. S., Naidoo, N., Teo, S. M., & Pawitan, Y. (2010). Regions of homozygosity and their impact on complex diseases and traits. *Hum Genet*, *129*(1), 1-15.

Lander, E. S. (2011). Initial impact of the sequencing of the human genome. *Nature*, *470*(7333), 187-197.

Lander, E. S., & Green, P. (1987). Construction of multilocus genetic linkage maps in humans. *Proceedings of the National Academy of Sciences*, *84*(8), 2363.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J. et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860-921.

Lapierre, L. A., & Goldenring, J. R. (2005). Interactions of myosin vb with rab11 family members and cargoes traversing the plasma membrane recycling system. *Methods Enzymol*, *403*, 715-723.

Lasky-Su, J. A., Faraone, S. V., Glatt, S. J., & Tsuang, M. T. (2005). Meta-analysis of the association between two polymorphisms in the serotonin transporter gene and affective disorders. *Am J Med Genet B Neuropsychiatr Genet*, *133B*(1), 110-115.

Laurie, C. C., Doheny, K. F., Mirel, D. B., Pugh, E. W., Bierut, L. J., Bhangale, T. et al. (2010). Quality control and quality assurance in genotypic data for genome-wide association studies. *Genetic epidemiology*, *34*(6), 591-602.

Lebel, R. R., & Gallagher, W. B. (1989). Wisconsin consanguinity studies. II: Familial adenocarcinomatosis. *American journal of medical genetics*, *33*(1), 1-6.

Lee, S. H., DeCandia, T. R., Ripke, S., & Yang…, J. (2012). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nature Genetics*.

Lencz, T., Lambert, C., DeRosse, P., Burdick, K. E., Morgan, T. V., Kane, J. M. et al. (2007). Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc Natl Acad Sci U S A*, *104*(50), 19942-19947.

Li, L. H., Ho, S. F., Chen, C. H., Wei, C. Y., Wong, W. C., Li, L. Y. et al. (2006). Long contiguous stretches of homozygosity in the human genome. *Hum Mutat*, *27*(11), 1115-1121.

Lindsay, A. J., & McCaffrey, M. W. (2009). Myosin Vb localises to nucleoli and associates with the RNA polymerase I transcription complex. *Cell Motil Cytoskeleton*, *66*(12), 1057-1072.

Lise, M. F., Wong, T. P., Trinh, A., Hines, R. M., Liu, L., Kang, R. et al. (2006). Involvement of myosin Vb in glutamate receptor trafficking. *J Biol Chem*, *281*(6), 3669-3678.

Liu, J., Aoki, M., Illa, I., Wu, C., Fardeau, M., Angelini, C. et al. (1998). Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi myopathy and limb girdle muscular dystrophy. *Nat Genet*, *20*(1), 31-36.

Maller, J., George, S., Purcell, S., Fagerness, J., Altshuler, D., Daly, M. J. et al. (2006). Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nature genetics*, *38*(9), 1055-1059.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J. et al. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*(7265), 747-753.

Mansour, H. A., Talkowski, M. E., Wood, J., Pless, L., Bamne, M., Chowdari, K. V. et al. (2005). Serotonin gene polymorphisms and bipolar I disorder: focus on the serotonin transporter. *Ann Med*, *37*(8), 590-602.

McCarroll, S. A., Hadnott, T. N., Perry, G. H., Sabeti, P. C., Zody, M. C., Barrett, J. C. et al. (2005). Common deletion polymorphisms in the human genome. *Nature genetics*, *38*(1), 86-92.

McGough, J. J., Loo, S. K., McCracken, J. T., Dang, J., Clark, S., Nelson, S. F. et al. (2008). CBCL Pediatric Bipolar Disorder Profile and ADHD: Comorbidity and Quantitative Trait Loci Analysis. *J Am Acad Child Adolesc Psychiatry*, *47*(10), 1151-1157.

McQueen, M. B., Devlin, B., Faraone, S. V., Nimgaonkar, V. L., Sklar, P., Smoller, J. W. et al. (2005a). Combined analysis from eleven linkage studies of bipolar disorder provides strong evidence of susceptibility loci on chromosomes 6q and 8q. *The American Journal of Human Genetics*, *77*(4), 582-595.

McQueen, M. B., Devlin, B., Faraone, S. V., Nimgaonkar, V. L., Sklar, P., Smoller, J. W. et al. (2005b). Combined analysis from eleven linkage studies of bipolar disorder provides strong evidence of susceptibility loci on chromosomes 6q and 8q. *Am J Hum Genet*, *77*(4), 582-595.

McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L. et al. (2008). Runs of homozygosity in European populations. *Am J Hum Genet*, *83*(3), 359-372.

Merikangas, K. R., Akiskal, H. S., Angst, J., Greenberg, P. E., Hirschfeld, R. M., Petukhova, M. et al. (2007). Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. *Arch Gen Psychiatry*, *64*(5), 543-552.

Momeni, P., Schymick, J., Jain, S., Cookson, M. R., Cairns, N. J., Greggio, E. et al. (2006). Analysis of IFT74 as a candidate gene for chromosome 9p-linked ALS-FTD. *BMC Neurol*, *6*, 44.

Morton, N. E. (1978). Effect of inbreeding on IQ and mental retardation. *Proceedings of the National Academy of Sciences*, *75*(8), 3906.

Muller, D. J., Serretti, A., Sicard, T., Tharmalingam, S., King, N., Artioli, P. et al. (2007). Further evidence of MAO-A gene variants associated with bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet*, *144B*(1), 37-40.

Muller, T., Hess, M. W., Schiefermeier, N., Pfaller, K., Ebner, H. L., Heinz-Erian, P. et al. (2008). MYO5B mutations cause microvillus inclusion disease and disrupt epithelial cell polarity. *Nat Genet*, *40*(10), 1163-1165.

Nalls, M. A., Guerreiro, R. J., Simon-Sanchez, J., Bras, J. T., Traynor, B. J., Gibbs, J. R. et al. (2009). Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. *Neurogenetics*, *10*(3), 183-190.

Need, A. C., Attix, D. K., McEvoy, J. M., Cirulli, E. T., Linney, K. L., Hunt, P. et al. (2009). A genome-wide study of common SNPs and CNVs in cognitive performance in the CANTAB. *Human molecular genetics*, *18*(23), 4650.

Neves-Pereira, M., Mundo, E., Muglia, P., King, N., Macciardi, F., & Kennedy, J. L. (2002). The brain-derived neurotrophic factor gene confers susceptibility to bipolar disorder: evidence from a family-based association study. *Am J Hum Genet*, *71*(3), 651-655.

Nothnagel, M., Lu, T. T., Kayser, M., & Krawczak, M. (2010). Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Hum Mol Genet*, *19*(15), 2927-2935.

Petronis, A. (2010). Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature*, *465*(7299), 721-727.

Pritchard, J. K. (2001). Are rare variants responsible for susceptibility to complex diseases? *The American Journal of Human Genetics*, *69*(1), 124-137.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D. et al. (2007a). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, *81*(3), 559-575.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D. et al. (2007b). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559-575.

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F. et al. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, *460*(7256), 748-752.

Rabbee, N., & Speed, T. P. (2006). A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics*, *22*(1), 7-12.

Rabbitt, P. M. A., McInnes, L., Diggle, P., Holland, F., Bent, N., Abson, V. et al. (2004). The University of Manchester longitudinal study of cognition in normal healthy old age, 1983 through 2003. *Aging Neuropsychology and Cognition*, *11*(2-3), 245-279.

Reich, D. E., & Lander, E. S. (2001). On the allelic spectrum of human disease. *TRENDS in Genetics*, *17*(9), 502-510.

Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J. et al. (2001). Linkage disequilibrium in the human genome. *Nature*, *411*(6834), 199-204.

Rhea, S. A., Gross, A. A., Haberstick, B. C., & Corley, R. P. (2006). Colorado twin registry. *Twin Research and Human Genetics*, *9*(6), 941-949.

Roeder, K., Devlin, B., & Wasserman, L. (2007). Improving power in genome-wide association studies: weights tip the scale. *Genet Epidemiol*, *31*(7), 741-747.

Roeder, K., Bacanu, S. A., Wasserman, L., & Devlin, B. (2006). Using linkage genome scans to improve power of association in genome scans. *The American Journal of Human Genetics*, *78*(2), 243-252.

Rudan, I., Skaric-Juric, T., Smolej-Narancic, N., Janicijevic, B., Rudan, D., & Klaric, I. M. (2004). Inbreeding and susceptibility to osteoporosis in Croatian island isolates. *Coll Antropol*, *28*(2), 585-601.

Rudan, I., Smolej-Narancic, N., Campbell, H., Carothers, A., Wright, A., Janicijevic, B. et al. (2003). Inbreeding and the genetic complexity of human hypertension. *Genetics*, *163*(3), 1011.

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C. et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, *449*(7164), 913-918.

Sachs, G. S., Thase, M. E., Otto, M. W., Bauer, M., Miklowitz, D., Wisniewski, S. R. et al. (2003). Rationale, design, and methods of the systematic treatment enhancement program for bipolar disorder (STEP-BD). *Biol Psychiatry*, *53*(11), 1028-1042.

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. et al. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, *230*(4732), 1350.

Schaffner, S. F., Foo, C., Gabriel, S., Reich, D., Daly, M. J., & Altshuler, D. (2005). Calibrating a coalescent simulation of human genome sequence variation. *Genome Res*, *15*(11), 1576-1583.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P. et al. (2004). Large-scale copy number polymorphism in the human genome. *Science*, *305*(5683), 525.

Sharp, A. J., Locke, D. P., McGrath, S. D., Cheng, Z., Bailey, J. A., Vallente, R. U. et al. (2005). Segmental duplications and copy-number variation in the human genome. *The American Journal of Human Genetics*, *77*(1), 78-88.

Shi, J., Levinson, D. F., Duan, J., Sanders, A. R., Zheng, Y., Pe'er, I. et al. (2009a). Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*, *460*(7256), 753-757.

Shi, J., Levinson, D. F., Duan, J., Sanders, A. R., Zheng, Y., & Peâ, I. (2009b). Common variants on chromosome 6p22. 1 are associated with schizophrenia. *Nature*, *460*(7256), 753-757.

Sklar, P., Smoller, J. W., Fan, J., Ferreira, M. A., Perlis, R. H., Chambert, K. et al. (2008). Whole-genome association study of bipolar disorder. *Mol Psychiatry*, *13*(6), 558-569.

Sklar, P., Gabriel, S. B., McInnis, M. G., Bennett, P., Lim, Y. M., Tsan, G. et al. (2002). Family-based association study of 76 candidate genes in bipolar disorder: BDNF is a potential risk locus. Brain-derived neutrophic factor. *Mol Psychiatry*, *7*(6), 579-593.

Smoller, J. W., & Finn, C. T. (2003). Family, twin, and adoption studies of bipolar disorder. *Am J Med Genet C Semin Med Genet*, *123C*(1), 48-58.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of molecular biology*, *98*(3), 503-517.

Spain, S. L., Cazier, J. B., Houlston, R., Carvajal-Carmona, L., & Tomlinson, I. (2009). Colorectal cancer risk is not associated with increased levels of homozygosity in a population from the United Kingdom. *Cancer research*, *69*(18), 7422.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(3), 479-498.

Sturtevant, A. H. (2001). *A history of genetics.* Cold Spring Harbor Laboratory Pr.

Sullivan, P. F., & Purcell, S. (2008). Analyzing genome-wide association study data: a tutorial using PLINK. In B. M. Neale, M. A. Ferreira, S. E. Medland, & D. Posthuma (Eds.), *Statistical genetics: gene mapping through linkage and association* (pp. 355-391). New York: Taylor & Francis Group.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G. et al. (2001). The sequence of the human genome. *science*, *291*(5507), 1304.

Vine, A. E., McQuillin, A., Bass, N. J., Pereira, A., Kandaswamy, R., Robinson, M. et al. (2009). No evidence for excess runs of homozygosity in bipolar disorder. *Psychiatr Genet*, *19*(4), 165-170.

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006a). A map of recent positive selection in the human genome. *PLoS biology*, *4*(3), e72.

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006b). A map of recent positive selection in the human genome. *PLoS Biol*, *4*(3), e72.

Walsh, T., Shahin, H., Elkan-Miller, T., Lee, M. K., Thornton, A. M., Roeb, W. et al. (2010). Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82. *Am J Hum Genet*, *87*(1), 90-94.

Wang, S., Haynes, C., Barany, F., & Ott, J. (2009). Genome-wide autozygosity mapping in human populations. *Genet Epidemiol*, *33*(2), 172-180.

Wechsler, D., & Corporation, P. (1991). *WISC-III: Wechsler intelligence scale for children.* The Psychological Corporation San Antonio.

Wechsler, D., & Corporation, P. (1997). *WAIS-III: Wechsler adult intelligence scale.* Psychological Corporation.

Weedon, M. N., Lango, H., Lindgren, C. M., Wallace, C., Evans, D. M., Mangino, M. et al. (2008). Genome-wide association analysis identifies 20 loci that influence adult height. *Nature genetics*, *40*(5), 575-583.

Whittemore, A. S., & Halpern, J. (1994). A class of tests for linkage using affected pedigree members. *Biometrics*, *50*(1), 118-127.

Williams, H. J., Norton, N., Dwyer, S., Moskvina, V., Nikolov, I., Carroll, L. et al. (2010). Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. *Mol Psychiatry*.

Woodley, M. A. (2009). Inbreeding depression and IQ in a study of 72 countries. *Intelligence*, *37*(3), 268-276.

Wright, M. J., & Martin, N. G. (2004). Brisbane adolescent twin study: outline of study methods and research projects. *Australian Journal of Psychology*, *56*(2), 65-78.

WTCCC. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, *447*(7145), 661-678.

Xiao, S., Sato, C., Kawarai, T., Goodall, E. F., Pall, H. S., Zinman, L. H. et al. (2008). Genetic studies of GRN and IFT74 in amyotrophic lateral sclerosis. *Neurobiol Aging*, *29*(8), 1279-1282.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R. et al. (2010a). Common SNPs explain a large proportion of the heritability for human height. *Nature genetics*, *42*(7), 565-569.

Yang, T. L., Guo, Y., Zhang, L. S., Tian, Q., Yan, H., Papasian, C. J. et al. (2010b). Runs of homozygosity identify a recessive locus 12q21.31 for human adult height. *J Clin Endocrinol Metab*, *95*(8), 3777-3782.

Zeggini, E., Scott, L. J., Saxena, R., Voight, B. F., Marchini, J. L., Hu, T. et al. (2008). Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature genetics*, *40*(5), 638-645.

Zelinger, L., Banin, E., Obolensky, A., Mizrahi-Meissonnier, L., Beryozkin, A., Bandah-Rozenfeld, D. et al. (2011). A missense mutation in DHDDS, encoding dehydrodolichyl diphosphate synthase, is associated with autosomal-recessive retinitis pigmentosa in Ashkenazi Jews. *Am J Hum Genet*, *88*(2), 207-215.

Zuk, O., Hechter, E., Sunyaev, S. R., & Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences*.